مجله علوم پزشکی رازی دوره ۲۶، شماره ۸، آبان ۱۳۹۸ http://rjms.iums.ac.ir

مقاله پژوهشی

ارائه مدلی جهت تشخیص ابتلا به بیماری نارسایی کلیوی با استفاده از تکنیک‌های یادگیری ماشین

سجاد باستین تختی: دانشجوی کارشناسی ارشد مهندسی صنایع، دانشگاه سیستان و بلوچستان، زاهدان، ایران

فرزاد فیروزی جهانتیغ: دکتری مهندسی صنایع، دانشیار، گروه مهندسی صنایع، دانشگاه سیستان و بلوچستان، زاهدان، ایران (*نویسنده مسئول). firouzi@eng.usb.ac.ir

چکیده

زمینه و هدف: امروزه کاربرد هوش مصنوعی درزمینه سیستم‌های سلامت گسترش زیادی داشته است. یادگیری ماشین به‌عنوان یکی از زیرشاخه‌های هوش مصنوعی، کاربردهای فراوانی درزمینه تشخیص پزشکی دارد. بیماری مزمن کلیوی یکی از شایع‌ترین بیماری‌های مربوط به کلیه در سراسر جهان است که تسهیل و تسریع در امر تشخیص آن نتایج بسیار مطلوبی بر روند درمان آتی آن خواهد داشت. هدف این پژوهش ارائه مدلی هوشمند برپایه‌ی تکنیک‌های یادگیری ماشین جهت تشخیص بیماری نارسایی کلیوی است.

روش کار: این مطالعه از نوع توصیفی-تحلیلی می‌باشد. داده‌های مورد استفاده در این تحقیق از تعداد ۴۰۰ فرد بیمار و غیر بیمار در کشور هندوستان استخراج‌شده است. این داده‌ها ابتدا در محیط پایتون پیش‌پردازش شده و از مشاهدات نویز و دورافتاده پاک شد. سپس الگوریتم‌های ماشین بردار پشتیبان، پرسپترون چندلایه و درخت تصمیم جهت دسته‌بندی داده‌ها به کار گرفته شد. معیارهای ارزیابی Accuracy، Recall و Precision برای ارزیابی عملکرد این دسته‌بندها محاسبه شد.

یافته‌ها: با توجه معیارهای ارزیابی محاسبه‌شده، برای الگوریتم ماشین بردار پشتیبان، مقادیر معیارهای Accuracy، Recall و Precision به ترتیب برابر ۰/۹۷، ۰/۹۶۱، ۰/۹۸۶ به دست آمد. یافته‌ها حکایت از عملکرد بهتر الگوریتم ماشین بردار پشتیبان ازنظر معیارAccuracy دارد. ازنظر معیار Recall، الگوریتم درخت تصمیم با مقدار ۰/۹۶۳ بهترین عملکرد را داشته و از نظر معیار Precision، الگوریتم پرسپترون چندلایه با مقدار ۰/۹۹۴ بهترین عملکرد را در دسته‌بندی داده‌ها داشتند.

نتیجه‌گیری: نتایج به‌دست‌آمده نشان داد که تکنیک‌های یادگیری ماشین می‌توانند در تشخیص بیماری مزمن کلیوی اثرگذار باشند. به‌کارگیری این تکنیک‌ها می‌تواند امور مربوط به تشخیص و درمان این بیماران را تسهیل کند و احتمال بهبودی افراد را بالا برد. همچنین نتایج نشان داد که مدل ارائه شده بر پایه‌ی تکنیک‌های یادگیری ماشین، در مقایسه با سایر تکنیک‌ها دقیق‌تر، ساده‌تر و کم‌هزینه تر است.

تعارض منافع: گزارش نشده است.

منبع حمایت کننده: حامی مالی نداشته است.

کلیدواژه‌ها

هوش مصنوعی،
یادگیری ماشین،
دسته‌بندی،
تشخیص پزشکی،
بیماری نارسایی کلیوی

تاریخ دریافت: ۹۸/۰۳/۱۹

تاریخ پذیرش: ۹۸/۰۷/۰۷

شیوه استناد به این مقاله:

Bastin Takhti S, Firouzi Jahantigh F. A model for diagnosis of kidney disease using machine learning techniques. Razi J Med Sci. 2019;26(8):14-22.

*انتشار این مقاله به‌صورت دسترسی آزاد مطابق با CC BY-NC-SA 1.0 صورت گرفته است.

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]

Razi Journal of Medical Sciences.2019;26(8):14-22. http://rjms.iums.ac.ir

Original Article

A model for diagnosis of kidney disease using machine learning techniques

Sajad Bastin Takhti, MSc Student in Industrial Engineering, University of Sistan and Baluchestan, Zahedan, Iran
Farzad Firouzi Jahantigh, PhD in Industrial Engineering, Associater Professor, University of Sistan and Baluchestan, Zahedan, Iran (*Corresponding author). firouzi@eng.usb.ac.ir

Abstract

Background: Today, the application of artificial intelligence in the field of health systems has been expanded. Machine learning as one of the sub-branches of artificial intelligence has many applications in the field of medical diagnosis. Chronic kidney disease is one of the most common kidney diseases around the world, which facilitation and acceleration in its diagnosis will have a very favorable outcome for its future treatment. The purpose of this study is to provide an intelligent model based on machine learning techniques for diagnosis of kidney diseases.

Methods: The data used in this study was extracted from 400 patients and non-patients in India. These data were pre-processed in the Python environment and cleared from noisy and outlying observations. Then support vector machine, multilayer perceptron, and decision tree were used for data classification. Accuracy, Recall and Precision evaluation metrics were calculated for performance evaluation of these classifiers.

Results: For the support vector machine algorithm, the Accuracy, Recall and Precision metrics were calculated to be 0.97, 0.961, and 0.986, respectively. The findings indicated that the support vector machine algorithm performs better in terms of Accuracy. In terms of Recall, the decision tree algorithm with the value of 0.963 had the best performance, and in terms of Precision, multi-layer perceptron algorithm with 0.994 had the best performance in data classification.

Conclusion: The results showed that machine learning techniques could be effective in the diagnosis of kidney disease. The use of these techniques can facilitate and expedite the diagnosis and treatment of these patients and increase the likelihood of recovery. The results also showed that the model presented on the basis of machine learning techniques is more accurate, simpler and less expensive than other techniques.

Conflicts of interest: None
Funding: None

Keywords
Artificial intelligence,
Machine learning,
Classification,
Medical diagnosis,
Kidney disease

Received: 09/06/2019
Accepted: 29/09/2019

Cite this article as:
Bastin Takhti S, Firouzi Jahantigh F. A model for diagnosis of kidney disease using machine learning techniques. Razi J Med Sci. 2019;26(8):14-22.

*This work is published under CC BY-NC-SA 1.0 licence.

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]

مقدمه

بیماری مزمن کلیوی به این معنی است که کلیه‌های انسان آسیب دیده‌اند و نمی‌توانند خون را آن‌طور که باید، پالایش کنند. این بیماری «مزمن» نامیده می‌شود، زیرا آسیب به کلیه‌های انسان به‌آرامی در طول مدت‌زمان طولانی اتفاق می‌افتد. این آسیب می‌تواند باعث ایجاد سموم در بدن شود. بیماری مزمن کلیوی همچنین می‌تواند مشکلات متعدد دیگری برای سلامتی ایجاد نماید. این بیماری حدود ۱۰ درصد از مردم دنیا را در برگرفته و شیوع آن نیز در حال افزایش است (۳-۱). در دنیای امروز با توجه به دسترسی فراوان به انواع داده‌های پزشکی می‌توان تکنیک‌های مختلف داده‌کاوی و هوش مصنوعی را برای تجزیه‌وتحلیل این داده‌ها به کار گرفت. این تجزیه‌وتحلیل‌ها می‌توانند به بهبود کیفیت خدمات پزشکی کمک شایانی نمایند. یادگیری ماشین به‌عنوان یکی از زیرمجموعه‌های علم هوش مصنوعی کاربردهای فراوانی در تشخیص و پیش‌بینی بروز انواع مختلف بیماری‌ها دارد (۴). تکنیک‌های یادگیری ماشین تاکنون در زمینه‌های مختلف پژوهش سلامت، مانند تشخیص انواع سرطان و دیابت به‌کار گرفته شده و نتایج مطلوبی از این بابت حاصل شده است (۷-۵). دونائوا و همکاران نیز با هدف تشخیص خودکار تومور، از روش‌های پردازش تصویر و یادگیری ماشین برای طبقه‌بندی تصاویر اندوسکوپی بهره جستند. آن‌ها توانستند با آموزش دادن به الگوریتم‌ها توسط تعداد کمی از این تصاویر به دقت بالایی در تصمیم‌گیری دست یابند (۸). در تحقیق دیگری، شین وو و همکاران با هدف تشخیص زودهنگام و پیشگیری از عواقب ناگوار بیماری کبد چرب، تکنیک‌های یادگیری ماشین را برای تشخیص آن به کار گرفتند. ایشان الگوریتم‌های جنگل تصادفی، شبکه‌های عصبی، نایو بیز و رگرسیون لجستیک را در تشخیص بیماری کبد چرب به کار بردند. معیارهای ارزیابی AUROC (The area under the receiver operating characteristic) و Accuracy نشان از کارایی نسبتاً خوب الگوریتم‌های یادشده در تشخیص بیماری کبد چرب داشتند (۹).

کاربرد تکنیک‌های یادگیری ماشین در امور مربوط به سلامت، محدود به بحث تشخیص پزشکی نمی‌شود. پیش‌بینی بقای بیماران یا پیش‌بینی بروز حالات مختلف در بیماران یا افراد تحت معالجه نیز از دیگر زمینه‌های کاربردی تکنیک‌های یادگیری ماشین است (۱۲-۱۰). شن و همکاران طی پژوهشی، از دو تکنیک مبتنی بر منطق فازی با عناوین FuRES (fuzzy rule-building expert system) و FOAM (fuzzy optimal associative memory) برای تشخیص بیماری مزمن کلیوی استفاده نمودند. آن‌ها همچنین از دسته‌بند خطی PLS-DA (partial linear squares discriminant analysis) جهت مقایسه با عملکرد دسته‌بندهای فازی استفاده کردند. مقادیر Accuracy به‌دست‌آمده برای تکنیک‌های FuRES و FOAM به ترتیب برابر با ۹۹,۲ و ۹۹٪ بود (۱۳). همچنین موتوکومار و کریشنان تکنیکی با عنوان intuitionistic fuzzy soft sets (IFSSs) را ارائه کردند. آنان کاربرد تکنیک پیشنهادی خود را در تشخیص پزشکی، با استفاده از معیارهای مختلفی ارزیابی نمودند. مقدار Accuracy برای تکنیک پیشنهادی ایشان در تشخیص بیماری مزمن کلیوی برابر با ۸۰٪ بود (۱۴). همان‌طور که پیش‌تر نیز اشاره شد استفاده از حجم عظیم داده‌های سیستم‌های سلامت در تشخیص پزشکی نتایج مطلوب و چشمگیری در پی داشته است. چنانچه یک مجموعه داده‌ای مفصل در اختیار باشد می‌توان با بکارگیری الگوریتم ها و تکنیک های مختلف یادگیری ماشین و یادگیری عمیق مدل هایی ایجاد نمود که با دقت بالایی بتوانند عمل طبقه بندی را روی مجموعه داده ای انجام دهند. این طبقه بندی می‌تواند بین افراد بیمار و غیر بیمار انجام شود که منجر به یادگیری یک مدل برای تشخیص بیماری می‌شود. این مدل‌ها می‌توانند به پزشکان در امر تشخیص بیماری‌ها یاری رسانده و خطاهای پزشکی ناشی از خستگی و ساعات کار طولانی را کاهش دهند. هدف تحقیق حاضر به‌کارگیری تکنیک‌های یادگیری ماشین درجهت کمک به تشخیص ابتلا به بیماری مزمن کلیوی و ارزیابی

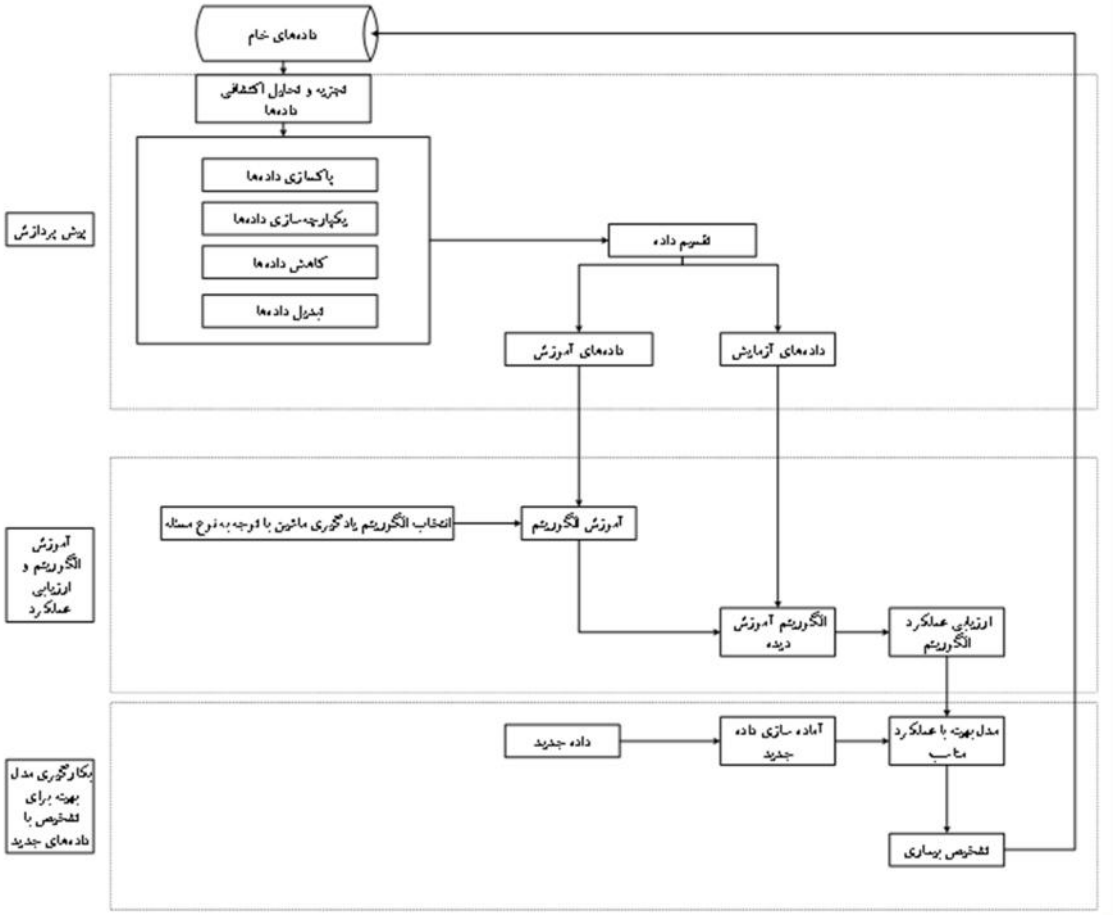
[Downloaded from rjms.iums.ac.ir on 2024-12-22]
[DOR: 20.1001.1.22287043.1398.26.8.7.1]

کارایی این تکنیک‌ها در مقابل سایر تکنیک‌های موجود است.

روش کار

این مطالعه از نوع توصیفی-تحلیلی می‌باشد. داده‌های استفاده شده در این مقاله از یک مرجع معتبر داده‌ای استخراج شده است. این داده‌ها مربوط به افرادی از کشور هندوستان می‌باشد که توسط نفرولوژیست ساندراپاندیان و همکاران وی گردآوری شده است و پژوهشگران این پژوهش مستقیماً نقشی در آن نداشته‌اند (۱۵). شمای کلی مدل تحقیق در شکل ۱ قابل ملاحظه است.

با توجه به شکل ۱، در مرحله پیش پردازش داده‌ها، ابتدا تجزیه و تحلیل اکتشافی داده‌ها صورت می‌پذیرد. تجزیه و تحلیل اکتشافی داده یک رویکرد برای تجزیه و تحلیل مجموعه داده ها به منظور ادراک ویژگی های اصلی آن هاست که می‌تواند با روش‌های بصری‌سازی همراه باشد (۱۶). سپس پاکسازی داده‌ها به منظور مدیریت مقادیر گمشده و مقادیر نویز انجام می‌شود. چنانچه داده‌ها از مراکز مختلف با فرمت‌ها و اشکال مختلف گردآوری شده باشند، لازم است یکپارچه‌سازی داده‌ها انجام شود. مجموعه داده مورد استفاده در پژوهش حاضر، شامل ۴۰۰ مشاهده (۲۵۰ فرد بیمار و ۱۵۰ فرد سالم) و ۲۵ مشخصه مربوط به این مشاهدات بود. اطلاعات تفصیلی مربوط به این مجموعه داده‌ای در جدول ۱ قابل مشاهده است. برای کاهش هزینه‌های محاسباتی و پردازشی چنانچه نیاز به استفاده از تمامی داده‌های در اختیار نباشد، در قسمت کاهش داده‌ها بخشی از داده‌ی مازاد بر نیاز کنار گذاشته می‌شود. تبدیل داده‌ها نیز شامل مواردی مثل نرمالسازی متغیرهای عددی و کدگزاری متغیرهای اسمی می‌باشد. همان‌طور که پیش‌تر نیز اشاره شد، یادگیری ماشین به‌عنوان یکی از زیرشاخه‌های هوش مصنوعی، کاربرد بسیار زیادی در مباحث مربوط به تشخیص پزشکی پیدا



شکل ۱- شمای کلی مدل تحقیق

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]

جدول ۱- مشخصات مجموعه‌ی داده‌ای

مشخصه	نوع مشخصه	مقادیر گمشده	میانگین در افراد مبتلا	میانگین در افراد غیر مبتلا
سن(سال)	عددی	۹	۵۴/۵۷	۴۶/۵۶
فشار خون(میلیمتر جیوه)	عددی	۱۲	۷۹/۷۹	۷۱/۲۹
وزن مخصوص	اسمی	۴۷	-	-
آلبومین	اسمی	۴۶	-	-
قند	اسمی	۴۹	-	-
گلبول قرمز	اسمی	۱۵۲	-	-
پوس سل	اسمی	۶۵	-	-
پوس سل کلامپس	اسمی	۴	-	-
باکتری	اسمی	۴	-	-
گلوکز(میلیگرم بر دسی لیتر)	عددی	۴۴	۱۷۵/۸۱	۱۰۷/۵۰
اوره(میلیگرم بر دسی لیتر)	عددی	۱۹	۷۲/۶	۳۲/۹۶
کراتینین(میلیگرم بر دسی لیتر)	عددی	۱۷	۴/۲۴	۰/۸۷
سدیم(میلی اکی والان بر لیتر)	عددی	۸۷	۱۳۳/۹۴	۱۴۱/۷۹
پتاسیم(میلی اکی والان بر لیتر)	عددی	۸۸	۴/۸۸	۴/۳۳
هموگلوبین(گرم بر لیتر)	عددی	۵۲	۱۰/۶۵	۱۵/۱۹
گلبول قرمز فشرده	عددی	۰	۳۲/۹۵	۴۶/۳۲
تعداد گلبول سفید(سلول بر میلیمتر مکعب)	عددی	۰	۹۰۶۴/۷۹	۷۶۸۶/۶۴
تعداد گلبول قرمز(میلیون سلول بر میلیمتر مکعب)	عددی	۰	۳/۹۷	۵/۳۹
هایپر تنشن	اسمی	۲	-	-
دیابت شیرین	اسمی	۲	-	-
بیماری قلبی عروقی	اسمی	۲	-	-
اشتها	اسمی	۱	-	-
ورم پا	اسمی	۱	-	-
آنمی	اسمی	۱	-	-
متغیر هدف (دسته)	اسمی	۰	-	-

کرده است. این‌گونه مسائل از دسته مسائل یادگیری ماشین با نظارت می‌باشند (Supervised learning). الگوریتم‌های مورد استفاده در این پژوهش باتوجه به نوع مسئله که از نوع یادگیری ماشین با نظارت می‌باشد، انتخاب شده‌اند. به‌طور مشخص ۸۰ درصد از داده‌ها به‌عنوان داده‌های آموزش و ۲۰ درصد جهت آزمایش الگوریتم‌ها استفاده شد. داده‌های آزمایش به منظور ارزیابی کارایی الگوریتم‌های یادگیری ماشین استفاده می‌شوند.

در این پژوهش ۳ مورد از پرکاربردترین الگوریتم‌های یادگیری ماشین، برای تشخیص بیماری مزمن کلیوی، با استفاده از مجموعه داده‌ای ذکر شده، به‌کار گرفته شده‌است. الگوریتم‌های استفاده‌شده در این پژوهش، درخت تصمیم، ماشین بردار پشتیبان و شبکه عصبی پرسپترون چندلایه می‌باشند.

درخت تصمیم: یکی از مهم‌ترین الگوریتم‌ها در حوزه مسائل مربوط به دسته‌بندی بوده و استفاده از آن به نسبت سایر الگوریتم‌ها آسان است. هر گره درخت تصمیم حاوی یک سؤال است که داده‌های عددی را بر اساس یک حد آستانه از پیش تعیین‌شده به گره‌های سطح پایین‌تر تخصیص می‌دهد. داده‌های اسمی نیز بر اساس یک سؤال و مقادیری که اختیار می‌کنند به گره‌های سطح پایین‌تر تخصیص می‌یابند. انتخاب بالاترین سطح گره‌ها یا گره ریشه (اولین گره و سؤال مطرح‌شده) موضوعی بسیار مهم است. روش‌های متعددی برای این انتخاب وجود دارد. شاخص بهره اطلاعاتی(Information gain) یکی مهم‌ترین روش‌ها برای انتخاب گره‌هاست (۱۷).

[Downloaded from rjms.iums.ac.ir on 2024-12-22]
[DOR: 20.1001.1.22287043.1398.26.8.7.1]

ماشین بردار پشتیبان: ماشین بردار پشتیبان نیز از الگوریتم‌های رایج و مهم در بحث طبقه‌بندی داده‌هاست. بردارهای پشتیبان به زبان ساده، مجموعه‌ای از نقاط در فضای n بعدی داده‌ها هستند که مرز دسته‌ها را مشخص می‌کنند و مرزبندی و دسته‌بندی داده‌ها بر اساس آن‌ها انجام می‌شود. الگوریتم ماشین بردار پشتیبان سعی می‌کند مرز بین دو دسته را طوری تعیین کند که بیشترین فاصله را از بردارهای پشتیبان هرکدام از دسته‌ها داشته باشد (۱۸).

شبکه عصبی پرسپترون چندلایه: یک نوع شبکه عصبی پرکاربرد جهت حل مسائل دسته‌بندی دودویی و چندگانه می‌باشد. شبکه‌های عصبی با الهام از سلول‌های عصبی مغز انسان طراحی شده‌اند. سلول‌های عصبی انسان یا همان نرون‌ها سیگنال‌های الکتریکی را به‌وسیله دندریت‌های خود دریافت نموده، پس از پردازش آن، سیگنالی از طریق آکسون خود، به نرون دیگر ارسال می‌کنند. عملکرد پرسپترون بر اساس همین عملکرد نرون طراحی شده است (۱۹).

برای ارزیابی عملکرد و کارایی این الگوریتم‌ها ابزار زیادی طراحی شده‌اند. در مسائل دسته‌بندی دوکلاسه از مواردی مثل Accuracy، Precision، Recall استفاده می‌شود. در روابط شماره ۱ تا ۴، T_p نشان‌دهنده‌ی تعداد تشخیص‌های صحیح گروه مثبت یا همان افراد بیمار است. T_n نشان‌دهنده‌ی تعداد تشخیص‌های گروه منفی یا همان افراد غیر بیمار است. F_p نشانگر تعداد افرادی است که به اشتباه بیمار تشخیص داده شده و F_n نشان‌گر تعداد افرادی است که به اشتباه غیر بیمار تشخیص داده شده‌اند. معیار Accuracy طبق رابطه‌ی [۱]، نشان دهنده‌ی نسبت تعداد پیش‌بینی‌های صحیح به کل پیش‌بینی‌ها می‌باشد. معیارهای Precision و Recall نیز از روابط [۲] و [۳] محاسبه می‌شوند (۲۰).

$$Accuracy = \frac{T_p + T_n}{ALL} \qquad [۱]$$

$$Precision = \frac{T_p}{T_p + F_p} \qquad [۲]$$

$$Recall = \frac{T_p}{T_p + F_n} \qquad [۳]$$

تمامی تجزیه و تحلیل، پیش پردازش و پردازش داده‌ها با استفاده پایتون ۲,۷ در ماژول ژوپیتر نوت بوک اجرا شد.

یافته‌ها

فاز اول این پژوهش مربوط به آماده‌سازی داده‌ها بود. دلیل آماده‌سازی یا پیش‌پردازش داده‌ها احتمال وجود برخی مشاهدات نویز و پرت یا وجود مقادیر گمشده می‌باشد. عدم آماده‌سازی مناسب داده‌ها تاثیر منفی در روند یادگیری الگوریتم‌ها دارد. به دلیل زیاد بودن مقادیر گمشده، حذف مشاهدات دارای مقادیر گمشده امکان‌پذیر نبود و بخش زیادی از مجموعه داده‌ای را از بین می‌برد. مشخصه‌ی گلبول قرمز طبق جدول ۱ بیشترین مقادیر گمشده را دارا بود. با توجه به تعداد بالای مقادیر گمشده در این مشخصه، به جهت اجتناب از بروز اشکلات احتمالی، این مشخصه از مجموعه‌ی داده‌ای کنار گذاشته شد. سایر مقادیر گمشده در داده‌ها در مورد مشخصه‌های عددی، با میانگین آن مشخصه و در مورد مشخصه‌های اسمی با مد آن مشخصه جایگزین شد. جهت از بین بردن مقیاس داده‌های عددی، تمامی مشخصه‌های عددی نرمال‌سازی شدند. عدم نرمالسازی داده‌ها با توجه به تفاوت مقیاس داده‌های ورودی ممکن است در روند آموزش الگوریتم‌ها اختلال ایجاد نماید. برای نرمالسازی داده‌ها از روش نرمال سازی آماری طبق رابطه [۴] استفاده شد. در این رابطه X نشان دهنده‌ی داده ورودی، μ_X نشان دهنده‌ی میانگین مشخصه X و σ_X نشان دهنده‌ی انحراف معیار مشخصه X است.

$$X_{norm} = \frac{X - \mu_X}{\sigma_X} \qquad [۴]$$

مقادیر داده‌های اسمی نیز با مقادیر عددی کدگزاری شد. برای مثال در مشخصه آنمی که نشان‌دهنده ابتلا یا عدم ابتلای فرد به آنمی می‌باشد، مقادیر بله و خیر به ترتیب با ۰ و ۱ جایگزین شد. به همین ترتیب سایر مشخصه‌های اسمی نیز کدگزاری شدند. پس از انجام بررسی‌های اولیه و آماده‌سازی داده‌ها، الگوریتم‌های یاد شده بر روی مجموعه داده‌ای به کار گرفته شد که نتایج حاصله به شرح جدول ۲ بود.

نتایج حاصل از جدول ۲ نشان می‌دهد که الگوریتم ماشین بردار پشتیبان از نظر معیار Accuracy با مقدار ۰,۹۷ یا به عبارتی ۹۷٪ تشخیص درست روی داده‌های آزمایش، بهترین عملکرد را داشت. از نظر معیار Recall، الگوریتم درخت تصمیم با مقدار ۰/۹۶۳ بهترین عملکرد را داشته و از نظر معیار Precision، الگوریتم پرسپترون

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]

جدول ۲- نتایج محاسبه معیارهای ارزیابی

	Accuracy	Recall	Precision
ماشین بردار پشتیبان	۰/۹۷	۰/۹۶۱	۰/۹۸۶
پرسپترون چندلایه	۰/۹۶۹	۰/۹۶۱	۰/۹۹۴
درخت تصمیم	۰/۹۵۸	۰/۹۶۳	۰/۹۸

چندلایه با مقدار ۰/۹۹۴ بهترین عملکرد را در دسته‌بندی داده‌ها داشتند. مهم‌ترین معیار محاسبه شده Accuracy می‌باشد زیرا بیانگر توانایی کلی یک الگوریتم در دسته‌بندی بین دو گروه بیمار و غیر بیمار می‌باشد که از این نظر می‌توان الگوریتم ماشین بردار پشتیبان را موفق‌تر از دو الگوریتم دیگر تلقی کرد.

بحث و نتیجه‌گیری

هدف از ارائه این پژوهش تسهیل و تسریع در تشخیص بیمار مزمن کلیوی و همچنین نشان دادن کارایی تکنیک‌های یادگیری ماشین در امر تشخیص پزشکی بود. در بحث توانایی تکنیک‌های یادگیری ماشین در تشخیص پزشکی می‌توان مقایسه‌ای بین نتایج حاصل از پژوهش‌های چن و موتوکومار (۱۳ و ۱۴) با تحقیق حاضر داشت. مقدار Accuracy به‌دست‌آمده از تحقیق حاضر برابر با ۹۷٪ برای الگوریتم ماشین بردار پشتیبان بود. این میزان از Accuracy در مقایسه با نتایج تحقیق (۱۳) که مقدار Accuracy آن برابر ۹۹٪ بود، تفاوت چندانی نداشته و مقدار ناچیزی کمتر می‌باشد؛ اما نسبت به تکنیک IFSSs به کار گرفته‌شده در (۱۴) که Accuracy برابر با ۸۰٪ دارد، بهبود قابل‌توجهی را نشان می‌دهد. همچنین مقادیر Accuracy به‌دست‌آمده از ۲ الگوریتم دیگر در تحقیق حاضر نیز به میزان قابل‌توجهی از Accuracy به‌دست‌آمده در (۱۴) بهتر بود. اگرچه نتایج حاصل از پژوهش‌های چن و موتوکومار (۱۳ و ۱۴)، نتایج نسبتاً مطلوبی است، اما قابلیت به‌کارگیری این تکنیک‌ها با توجه به مباحث تئوری پیچیده و عدم وجود تسهیلات نرم‌افزاری خاص این تکنیک‌ها، جای بحث دارد. شیخ‌طاهری و همکاران نیز با استفاده از منطق فازی، یک سیستم خبره فازی برای تشخیص بیماری مزمن کلیوی ارائه نمودند. ایشان سیستم خبره پیشنهادی خود را یا استفاده از سیستم استنتاج ممدانی طراحی نموده و عملکرد این سیستم را بااستفاده از داده‌های ۲۱۶ فرد بیمار و غیر بیمار مورد ارزیابی قرار دادند. معیار Accuracy به دست‌آمده از پژوهش شیخ‌طاهری و همکاران حدود ۹۰٪ بود. ایشان کمبود داده‌های پزشکی مرتبط را از محدودیت‌های پژوهش خویش ذکر نموده‌اند. در ذیل پژوهش ایشان ذکر شده که نتایج حاصل از تکنیک‌های یادگیری ماشین بهتر از مدل پیشنهادی ایشان بوده‌است. اما در عین حال داده‌های مورد استفاده در پژوهش خود را بسیار حائز اهمیت‌تر و مرتبط‌تر از پژوهش حاضر دانسته‌اند (۲۱). یکی از اساسی‌ترین مزایای تکنیک‌های یادگیری ماشین، وجود تسهیلات نرم‌افزاری قدرتمند جهت به‌کارگیری آن‌ها است. وجود کتابخانه‌های قدرتمند در زبان برنامه‌نویسی پایتون، کمک شایانی به عملیاتی شدن و استفاده از مباحث یادگیری ماشین نموده است. چنانچه رویکردی عملیاتی در هوشمندسازی روند تشخیص بیماری مزمن کلیوی مدنظر باشد، گزینه‌ی ارجح استفاده از تکنیک‌های یادگیری ماشین است. در پژوهش آکبن (۲۲)، الگوریتم یادگیری ماشین KNN(k nearest neighbor) در حالتی که از تمامی مشخصه‌های مجموعه‌ی داده‌ای برای آموزش الگوریتم استفاده می‌نمود، توانست به Accuracy معادل ۹۶٪ دست یابد. در ادامه با سیری پیچیده از پیش‌پردازش داده‌ها و استفاده از روش‌های خوشه‌بندی، الگوریتم KNN توانست میزان Accuracy را به عددی بین ۹۶ تا ۹۸٪ برساند. تحقیق حاضر بدون استفاده از تکنیک‌های سخت و پیچیده در پیش‌پردازش داده‌ها و صرفاً با استفاده از تکنیک‌های ساده به Accuracy معادل پژوهش آکبن (۲۲) رسید. سینها و همکار وی، در پژوهشی، دو تکنیک یادگیری ماشین KNN و ماشین بردار پشتیبان را برای تشخیص بیماری مزمن کلیوی به‌کار بردند. در پژوهش ایشان الگوریتم KNN عملکرد بهتری از خود نشان داد اما به‌طورکلی مقدار Accuracy

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]

به‌دست‌آمده از این پژوهش بسیار پایین و زیر ۸۰٪ بود (۲۳). از این نظر نتایج پژوهش سینها و همکار وی همخوانی چندانی با نتایج پژوهش حاضر و پژوهش‌های بررسی شده ندارد. نتایج نسبتاً ضعیف پژوهش سینها (۲۳) می‌تواند ناشی از عواملی همچون عدم پیش‌پردازش مناسب داده‌ها باشد. نتایج حاصل از پژوهش حاضر نشان از کارایی بسیار مطلوب تکنیک‌های یادگیری ماشین در بحث تشخیص پزشکی دارد. این نتایج به‌نوعی با نتایج پژوهش‌های مرکالدو (۶)، دناوا (۷) ، وو (۹) و هروی (۲۴) همخوانی دارد، چراکه در آن پژوهش‌ها نیز کارایی تکنیک‌های یادگیری ماشین در بحث تشخیص پزشکی نشان داده شده است.

با توجه به معیارهای ارزیابی حاصله و بحث‌های مطرح‌شده، مشخص شد، تکنیک‌های یادگیری ماشین استفاده شده در پژوهش حاضر علاوه بر توانایی مطلوب در دسته‌بندی بین افراد مبتلا و غیر مبتلا به بیماری مزمن کلیوی، بار پردازشی کمتر در مقایسه با چن (۱۳)، موتوکومار (۱۴) و آکبن (۲۲) و درنتیجه قابلیت به‌کارگیری بیشتری دارند. تمامی تکنیک‌های به‌کاربرده شده در این پژوهش توانستند با دقتی بالای ۹۵٪ بیماری نارسایی کلیوی را روی مشاهدات جدید تشخیص دهند و این مقدار از دقت می‌تواند طراحان سیستم‌های سلامت را متقاعد به به‌کارگیری این تکنیک‌ها در پیکره سیستم برای کاهش خطا در تشخیص پزشکی نماید. با توجه سهولت به‌کارگیری تکنیک‌های یادگیری ماشین نسبت به سایر تکنیک‌های مطرح شده، داده‌های فراوان پزشکی و توانمندی رایانه‌های امروزی موجود، پیشنهاد می‌گردد سیستمی عملیاتی مبتنی بر هوش مصنوعی یا به طور خاص تکنیک‌های یادگیری ماشین، در مراکز درمانی کشور، جهت تشخیص بیماری مزمن کلیوی طراحی شود. همچنین پیشنهاد می‌شود این سیستم طراحی شده، از جنبه‌ی تعامل و سازگاری با HIS(health information system) یا سیستم‌های اطلاعات سلامت مورد بررسی قرار گیرد.

یکی از محدودیت‌های این پژوهش تعداد کم داده‌های این پژوهش و همچنین ایرانی نبودن بیماران موجود در این مجموعه‌ی داده‌ای بود. لذا، ممکن است الگوریتم‌های آموزش داده‌شده، در تشخیص بیماران ایرانی، به لحاظ احتمال تفاوت در برخی مشخصه‌ها نتایج خوبی ارائه ندهند. در نتیجه طراحی یک بانک داده‌ای الکترونیک برای این بیماران در مراکز درمانی و غربالگری، گامی مهم در جهت فراهم‌سازی داده‌های ارزشمند و ساختاریافته جهت کاربردهای هوشمندسازی در تشخیص این بیماری خواهد بود.

References

1. Kimmel P, Rosenberg M. Chronic Renal Disease. 1st ed. USA: Academic Press; 2014.
2. Pérez-Sáez MJ, Prieto-Alhambra D, Barrios C, Crespo M, Redondo D, Nogués X, et al. Increased hip fracture and mortality in kidney disease individuals: the importance of competing risks. Bone; 2015.73:154-9.
3. Cueto-Manzano AM, Cortés-Sanabria L, Martínez-Ramírez HR, Rojas-Campos E, Gómez-Navarro B, Castillero-Manzano M. Prevalence of kidney disease in an adult population. Arch Med Res; 2014.45(6):507-13.
4. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J; 2017.15:104-16.
5. Gharaati Z, Pajoohan M. [Diagnosis of leukemia type by machine learning: dimension reduction and balancing]. IJDO; 2018.5(1):25-34. [Persian]
6. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform; 2017.97:120-7.
7. Mercaldo F, Nardone V, Santone A. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. KES; 2017.112:2519-28.
8. Dunaeva O, Edelsbrunner H, Lukyanov A, Machin M, Malkova D, Kuvaev R, et al. The classification of endoscopy images with persistent homology. Patrec; 2016.83:13-22.
9. 9- Wu CC, Yeh WC, Hsu WD, Islam MM, Nguyen PA, Poly TN, et al. Prediction of fatty liver disease using machine learning algorithms. Comput Methods Programs Biomed; 2019.170:23-9.
10. Lynch C, Abdollahi B, Fuqua J, Carlo A, Bartholomai J, Balgemann R, et al. Prediction of lung cancer patient survival via supervised machine learning classification techniques. Int J Med Inform; 2017.108:1-8.
11. Dumortier A, Beckjord E, Shiffman S, Sejdić E. Classifying smoking urges via machine learning. Comput Methods Programs Biomed; 2016.137:203-13.

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]

12. Martínez-Martínez JM, Escandell-Montero P, Barbieri C, Soria-Olivas E, Mari F, Martínez-Sober M .et al. Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques. Comput Methods Programs Biomed; 2014.117(2):208-17.
13. Chen Z, Zhang Z, Zhu R, Xiang Y, Harrington PB. Diagnosis of patients with kidney disease by using two fuzzy classifiers. CMBEBIH; 2016.153:140-5.
14. Muthukumar P, Krishnan GS. A similarity measure of intuitionistic fuzzy soft sets and its application in medical diagnosis. Appl Soft Comput; 2016.41:148-56.
15. UCI Machine Learning Repository: Kidney disease (CKD) Data Set, 2015. [Internet]. Available from: https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease
16. Downey AB. Think stats: exploratory data analysis. USA: O'Reilly Media, Inc.; 2014.
17. Cady F. The Data Science Handbook. 1st ed. USA: John Wiley & Sons; 2017.
18. Deng N, Tian Y, Zhang C. Support vector machines: optimization based theory, algorithms, and extensions. 1st ed. USA: Chapman and Hall/CRC; 2012.
19. Soares FM, Souza AM. Neural network programming with Java. 1st ed: Birmingham UK: Packet Publishing Ltd; 2017.
20. Hackeling G. Mastering Machine Learning with scikit-learn. 1st ed: Birmingham UK: Packet Publishing Ltd; 2017.
21. Sheikhtaheri A, Hamedan F, Sanadgol H, Orooji A. [Development of a fuzzy expert system to diagnose kidney disease]. Razi J Med Sci; 2019.25(10):46-60. [Persian]
22. Akben SB. Early stage kidney disease diagnosis by applying data mining methods to urinalysis, blood analysis and disease history. IRBM; 2018.39(5):353-8.
23. Sinha P, Sinha P. Comparative study of kidney disease prediction using KNN and SVM. Int J Eng Res Technol; 2015.4:608-12.
24. Heravi M, Setayeshi S. [Intelligent and fast recognition of heart disease based on synergy of linear neural network and logistic regression model]. J Mashhad Uni Med Sci; 2014.24(112):78-87. [Persian]

[Downloaded from rjms.iums.ac.ir on 2024-12-22]

[DOR: 20.1001.1.22287043.1398.26.8.7.1]