پیش‌بینی و تعیین عوامل مؤثر بر بقای پنج‌سالهٔ کلیهٔ پیوندی در داده‌های نامتوازن با رویکرد فراابتکاری و یادگیری ماشین

نسیبه امامی* **و زینب حسنی**

گروه علوم کامپیوتر، دانشکده علوم پایه، دانشگاه کوثر بجنورد، ایران

چکیده

در مرحلهٔ نهاییِ نارسایی کلیه، پیوند کلیه می‌تواند عمر بیماران را طولانی کند و کیفیت زندگی بیمار را بسیار بهبود بخشد. بعد از عمل پیوند کلیه، بررسی میزان یا پیش‌بینی بقای کلیهٔ پیوندی اهمیت زیادی دارد. این مطالعه بر روی بیماران کلیهٔ پیوندی بیمارستان‌های امام رضا(ع) و چهارمین شهید محراب کرمانشاه در سال‌های ۲۰۱۲- ۲۰۰۱ انجام شده است. از آن‌جایی‌که داده‌های نامتوازن باعث ناکارامدی مدل‌های یادگیری ماشین می‌شوند، ابتدا داده‌های نامتوازن با دو روش بیش‌نمونه‌برداری و زیرنمونه‌برداری متوازن شدند؛ سپس عوامل اثرگذار بر بقای پیوند کلیه به‌کمک الگوریتم فراابتکاری ژنتیک شناسایی شده و مدل یادگیر طبقه‌بند نزدیک‌ترین همسایه برای پیش‌بینی بقای پنج سالهٔ کلیهٔ پیوندی به‌کار گرفته شد. بقای کلیهٔ پیوندی در روش بیش‌نمونه‌برداری با دقّت ۹۶/۸ درصد و زیرنمونه‌برداری با دقّت ۸۹/۲ درصد پیش‌بینی شد. هم‌چنین، ویژگی‌های وزن، سنِّ دهنده و گیرنده، اورهٔ قبل پیوند، کراتین قبل پیوند، هموگلوبین قبل و بعد پیوند، جنسیتِ دهنده، RH دهنده و گیرنده، بیماری اولیه، سنِّ دهندهٔ بالای سی و سنِّ گیرندهٔ بالای چهل، به‌عنوان ویژگی‌های تأثیرگذار در بقای کلیهٔ پیوندی شناسایی شد. مقایسهٔ نتایج به‌دست‌آمده از این پژوهش با مطالعات پیشین، برتری مدل پیشنهادی را از نقطه‌نظر دقّت مدل نشان می‌دهد. به‌عبارتی متوازن‌سازی داده‌ها همراه با انتخاب ویژگی بهینه منجر به ارائهٔ مدل پیش‌بینی دقیق‌تری می‌شود.

واژگان کلیدی: پیوند کلیه، داده‌های نامتوازن، الگوریتم ژنتیک، نزدیک‌ترین همسایگی.

Prediction and determining the effective factors on the survival transplanted kidney for five-year in imbalanced data by the meta-heuristic approach and machine learning

Nasibeh Emami* & Zeinab Hassani

Department of Basic Science, Kosar University of Bojnord, Iran

Abstract

Chronic kidney failure is one of the most widespread diseases in Iran and the world. In general, the disease is common in high health indexes societies due to increased longevity. Treatment for chronic kidney failure is dialysis and kidney transplantation. Kidney transplantation is an appropriate and effective strategy for patients with End-Stage Renal Disease (ESRD), and it provides a better life and reduces mortality risk for patients. In contrast to many benefits that kidney transplantation has in terms of improving physical and mental health and the life's quality in kidney transplantation patients, it may be rejected because of host's immune response to the received kidney, and it consequences the need for another transplantation, or even death will have to. In fact, a patient that can survive for years with dialysis, he may lose his life with an inappropriate transplantation or be forced into high-risk surgical procedures.

According to the above, the study of predicting the survival of kidney transplantation, its effective factors and providing a model for purposing of high prediction accuracy is essential. Studies in the field of survival

* Corresponding author

*نویسندهٔ عهده‌دار مکاتبات • تاریخ ارسال مقاله: ۱۳۹۶/۱۰/۴ • تاریخ آخرین بازنگری: ۱۳۹۷/۵/۱۶ • تاریخ پذیرش: ۱۳۹۷/۷/۱۴

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.29252/jsdp.15.4.85]

of kidney transplantation include statistical studies, artificial intelligence and machine learning. In all of the studies in this feild, researchers have sought to identify a more effective set of features in survival of transplantation and the design of predictive models with higher accuracy and lower error rate.
This study carried out on 756 kidney transplant patients with 21 features of Imam Reza and Fourth Shahid Merab hospital in Kermanshah from 2001 to 2012. Some features set to binary value and other features have real continuous values. Due to data are unbalance, which led to convergence of classification model to majority class, so over sampling and under sampling techniques has been used for achieving higher accuracy.
To identify the more effective features on the survival of the kidney transplantation, the genetic meta-heuristic algorithm is used. For this purpose binary coding for each chromosome has been used; it is combining three single-point, two-point, and uniform operators to make better generations, better convergence and achieve higher accuracy rate. The genetic search algorithm plays a vital role in searching for such a space in a reasonable time because data search space is exponential. In fact, in balanced data, genetic algorithm determines the effective factors and the K-nearest neighbor model with precision of classification as the evaluator function was used to predict the five-year survival of the kidney transplantation. Based on the results of this study, in comparison to similar studies for prediction of survival transplanted kidney, the five-year survival rate of transplanted kidney was appropriate in these models. Also the effective factors in over sampling and under sampling methods with a precision of 96.8% and 89.2% are obtained respectively. in addition weight, donor and recipient age, pre-transplantation urea, pre-transplantation creatinine, hemoglobin before and after transplantation, donor gender, donor and recipient RH, primary illness, donor age up 30 and receipt age up 40 were identified as the effective features on kidney transplantation survival. Comparing the results of this study with previous studies shows the superiority of the proposed model from the point of view of the models' precision. In particular, balancing the data along the selection of optimal features leads to a high precision predictive model.

Keywords: Kidney Transplantation, imbalance data, Genetic Algorithm, K- nearest neighbors.

۱- مقدمه

یکی از مهم‌ترین بیماری‌های کلیوی، نارسایی کلیه است که با کاهش شدید عملکرد کلیه همراه است. مهم‌ترین اقدام برای درمان بیماری، برطرف‌کردن عامل بیماری است تا باقی‌مانده بافت سالم کلیه حفظ شود و کلیه‌ها بتوانند کار خود را به‌خوبی انجام دهند. در مراحل اولیه نارسایی کلیهٔ بیمار با مصرف دارو ممکن است، بیماری‌اش را کنترل کند، اما اگر نارسایی کلیه در مراحل پیشرفته باشد، کلیه قادر نخواهد بود عمل تصفیه خون را به‌خوبی انجام دهد و سموم را از بدن خارج کند. بنابراین افراد مبتلا باید با روش مصنوعی دیالیز و درنهایت، پیوند کلیه درمان شوند. در مرحلهٔ نهاییِ نارسایی کلیه، پیوند کلیه، زندگی مطلوب‌تر و کاهش خطر مرگ و میر را برای بیماران فراهم می‌کند [1]. سابقهٔ پیوند کلیه به سال ۱۹۵۴ میلادی (۱۳۳۳شمسی) در آلمان بر می‌گردد. نخستین پیوند کلیه در ایران در سال ۱۹۶۷ در شیراز انجام گرفت. منابع تأمین عضو پیوندی شامل افراد زندهٔ فامیل، افراد زندهٔ غیرفامیل و جسد است. یکی از اهداف اصلی برنامه‌های پیوند عضو، فراهم‌کردن عضو پیوندی مناسب برای هر بیمار نیازمند به آن است که امروزه کمبود عضو، بزرگ‌ترین مانع در دست‌یابی به این هدف به‌شمار می‌رود [2]. بررسی میزان یا پیش‌بینی بقای کلیهٔ پیوندی بعد از عمل پیوند اهمیت زیادی دارد؛ زیرا بیماری که می‌تواند سال‌ها با دیالیز زندگی کند با پیوندی نامناسب ممکن است، زندگی خود را از دست دهد؛ بنابراین انجام مطالعه بر روی پیش‌بینی بقای کلیهٔ پیوندی و عوامل مؤثر بر آن و ارائهٔ یک مدل به‌منظور دقّت پیش‌بینی بالاتر ضرورت می‌یابد و در مراکز متعدد مطالعات مرتبط با بقای کلیهٔ پیوندی انجام گرفته است [3,1]. مطالعات انجام‌شده جهت بررسی بقای پیوند از نوع مطالعات آماری و هوش مصنوعی است [1]. تجزیه و تحلیل بقا را با استفاده از استانداردهای آماری، می‌توان مدل مبتنی بر جمعیت در نظر گرفت که پیش‌بینی‌ها براساس احتمال و یا فاصله از تخمین‌های جمعیت مشتق می‌شود. یادگیری ماشین و داده‌کاوی، ابزارهای تصمیم‌گیری را به‌جای یک جمعیت از بیماران برای هر بیمار ارائه می‌دهد که این وجه تمایز آن است. در اینجا تأکید بر یافتن دانش مفید، معتبر و کاملی از داده‌هاست [5]. هدف از این مطالعه، شناسایی عوامل مؤثر و برآورد میزان بقای پنج‌سالهٔ کلیهٔ پیوندی به‌کمک الگوریتم ژنتیک و روش‌های یادگیری ماشین و داده‌کاوی است که بر روی داده‌های بیمارستان امام رضا(ع) و چهارمین شهید محراب کرمانشاه در سال‌های ۲۰۱۲-۲۰۰۱ انجام شده است. با توجه به این‌که داده‌های مورد استفاده در این مطالعه نامتوازن است و داده‌های نامتوازن باعث ناکارامدی نتایج می‌شوند؛ از این‌رو برای بهبود نتایج متوازن‌سازی داده‌ها ضرورت دارد.

این مطالعه در ۶ بخش سازمان‌دهی شده است. در بخش۲ به پیشینهٔ پژوهش و در بخش ۳ به مواد و روش‌ها خواهیم پرداخت؛ در بخش ۴ به بیان الگوریتم پیشنهادی می‌پردازیم؛ سپس یافته‌ها و بحث در بخش ۵ مطرح می‌شود و در بخش ۶ نتیجه‌گیری آمده است.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.29252/jsdp.15.4.85]

۲- پیشینهٔ پژوهش

در سال ۲۰۰۷ سنتوری و همکاران به ارائهٔ مدل یادگیری شبکهٔ عصبی برای پیش‌بینی نتیجه بقای کلیهٔ پیوندی در مورد پیوند کلیه اطفال پرداخته‌اند که دقّت مدل حاصل ۸۷/۱۴ درصد است [6]. در سال ۲۰۰۹، اشرفی و همکاران جهت بررسی بقای کلیهٔ پیوندی در اصفهان با استفاده از شبکهٔ عصبی مدلی با دقّت ۹۰ درصد ارائه داده‌اند. هم‌چنین نشان دادند بر اساس نتایج این مطالعه، میزان بقای کلیهٔ پیوندی در پیوندهای زنده، بیش‌تر از پیوندهای جسدی است [3].

در مطالعهٔ انجام‌شده در سال ۲۰۱۰ در ایران توسط حشیانی و همکاران، عامل سنّ دهنده و سنّ گیرنده عامل معناداری بر بقای کلیهٔ پیوندی شناسایی شدند [4]. در سال ۲۰۱۲ در آمریکا، برای مدل‌سازی ۵۱۴۴ مورد پیوند کلیه با ۴۸ متغیر از روش شبکهٔ بیزی استفاده شد، شاخص تودهٔ بدنی، نژاد، جنسیت گیرنده و سنّ دهنده از عوامل مؤثر در بقای پیوند شناسایی شدند. در مطالعه‌ای در سال ۲۰۱۳ در آلمان برای پیش‌بینی میزان فیلتراسیون گلومار گیرنده یک سال پس از پیوند، روش ماشین‌بردار پشتیبان استفاده شد و متغیرهای مؤثر سن و سطح کراتنین دهنده و جنسیت و وزن گیرنده شناسایی شدند [8]. در سال ۲۰۱۳ بیرانوند و همکاران با ارائهٔ مدل شبکهٔ عصبی پرسپترون چندلایه بر روی ۷۵۶ دادهٔ بیمارستان امام رضا(ع) و چهارمین شهید محراب کرمانشاه که در سال‌های ۲۰۱۲-۲۰۰۲ جمع‌آوری شده است به پیش‌بینی بقای پنج سالهٔ کلیهٔ پیوندی پرداخته و مدلی با دقّت ۸۱ درصد به‌دست آورده‌اند [9]. میرزائی و همکاران در سال ۲۰۱۶، مطالعه‌ای بر روی اطلاعات پروندهٔ ۴۲۳ بیمار پیوند کلیهٔ مرکز آموزشی- درمانی افضلی‌پور شهر کرمان انجام داده‌اند. آنها الگوریتم ژنتیک را جهت شناسایی متغیرهای تأثیرگذار به‌کار برده‌اند. ویژگی‌های شاخص توده بدنی، جنسیت گیرنده، سنّ دهنده، همسانی گروه خونی دهنده و گیرنده و سابقهٔ پیوند کلیه به‌عنوان متغیرهای بهینه شناسایی شدند که صحت پیش‌بینی مدل پیشنهادی با این تعداد متغیر ۹۱،۶۷ درصد است [1].

در تمامی مطالعات انجام‌شده در این زمینه، پژوهش‌گران به‌دنبال شناسایی مجموعه‌ویژگی‌های عوامل مؤثرتر در بقای پیوند و هم‌چنین طراحی مدل‌های پیش‌بینی با دقّت بالاتر و خطای پایین‌تر بوده‌اند.

به‌طورعمومی داده‌های پزشکی، داده‌های نامتوازن هستند؛ به عبارتی در مجموعه داده‌های پزشکی، تعداد نمونه‌های یک طبقه به نام طبقهٔ بیشینه بیشتر از طبقهٔ دیگر به نام طبقهٔ کمینه هستند. تعداد نمونه داده‌های بیماران بیشتر از تعداد نمونه‌های افراد سالم است. الگوریتم‌های طبقه‌بندی با داده‌های نامتوازن باعث هم‌گراشدن طبقهٔ کمینه به طبقهٔ بیشینه می‌شود؛ به‌طوری‌که نمونه‌های طبقهٔ کمینه در طبقهٔ بیشینه طبقه‌بندی می‌شوند؛ درنهایت طبقه‌بندی ناکارا خواهد بود. یکی از روش‌های حل این مشکل، استفاده از روش‌های نمونه‌سازی از جمله بیش‌نمونه‌برداری و زیرنمونه‌برداری است. اسکویی و همکارانش، در سال ۲۰۱۵ دو روش نمونه‌برداری (بیش‌نمونه‌برداری و زیرنمونه‌برداری) را با سیزده مجموعه داده‌های نامتوازن با الگوریتم‌های طبقه‌بندی بیزین ساده و J48 مطالعه کرده‌اند و نشان داده‌اند که روش بیش‌نمونه‌برداری در الگوریتم طبقه‌بندی دارای دقّت بالاتری است [10]. در این مطالعه سعی بر ارائهٔ مدلی است که دقّت بالاتری نسبت به مدل‌های پیشین داشته باشد.

۳- مواد و روش‌ها

در این بخش به بیان مقدماتی راجع به مواد و روش‌های به‌کاررفته در مدل پیشنهادی می‌پردازیم.

۳-۱- نمونه‌برداری

در مجموعه‌داده‌های متوازن، تعداد نمونه‌ها در هر طبقه مجموعه‌داده برابر است؛ درصد متوازن‌سازی مجموعه داده‌های متوازن، در حدود پنجاه درصد است. درحالی‌که در مجموعه داده‌های نامتوازن، تعداد نمونه‌های یک طبقه از طبقهٔ دیگر بیشتر است؛ درصد متوازن‌سازی مجموعه داده‌های نامتوازن، خیلی کمتر از پنجاه درصد است. از طرفی طبقه‌بندی داده‌های نامتوازن ناکارا است. بنابراین طبقه‌بندی مجموعه‌داده‌های نامتوازن یک چالش مهم در یادگیری ماشین است [10].

در الگوریتم‌های استاندارد حوزهٔ یادگیری ماشین و داده‌کاوی، فرض بر این است که توزیع طبقه‌ها متوازن باشد. از این‌رو درصورت استفاده از این الگوریتم‌ها در طبقه‌بندی داده‌های نامتوازن، مدل به‌دست‌آمده به سمت نمونه‌های آموزشی طبقهٔ بزرگ‌تر متمایل می‌شود که سبب کاهش دقّت مدل حاصل در پیش‌بینی طبقهٔ کمینه می‌شود [12,11]؛ لذا در این مطالعه جهت رسیدن به مدلی با دقّت بهتر در مرحلهٔ پیش‌پردازش‌داده، از روش نمونه‌برداری برای متوازن‌سازی داده‌ها استفاده می‌شود. نمونه‌برداری به دو روش بیش‌نمونه‌برداری از طبقهٔ کمینه و زیرنمونه‌برداری از طبقهٔ بیشینه است.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.29252/jsdp.15.4.85]

۱-۱-۳- بیش‌نمونه‌برداری

این روش برای تعادل‌رساندن توزیع طبقه، از روش جایگزینی نمونه‌های طبقهٔ کمینه استفاده می‌کند، تا توازن در مجموعه آموزشی برقرار شود.

۲-۱-۳- زیرنمونه‌برداری

در این روش به‌طور تصادفی نمونه‌هایی از طبقهٔ بیشینه حذف می‌شود؛ تا زمانی که طبقهٔ کمینه درصدی از طبقهٔ بیشینه شود. به این ترتیب توازن در مجموعهٔ آموزشی برقرار می‌شود [14,13].

۲-۳- الگوریتم ژنتیک

الگوریتم ژنتیک در سال ۱۹۷۵ توسط هلند بر طبق نظریه تکاملی داروین ارائه شده است. الگوریتم با یک مجموعه از جمعیت (جواب‌های اولیه) که از طریق کروموزوم نشان داده می‌شوند، شروع می‌شود. در الگوریتم ژنتیک هر کروموزم به‌عنوان راه حل مسئله در نظر گرفته می‌شود. به‌طوری‌که هر کروموزوم شامل تعدادی ژن است که به‌صورت عددی برطبق مسئله کدگذاری شده است. تعداد جمعیت در الگوریتم ژنتیک براساس پیچیدگی مسئله تعیین می‌شود. ابتدا کروموزوم‌ها براساس مسئله کدگذاری و در ادامه پارامترهای الگوریتم از جمله تعداد جمعیت تعیین می‌شود؛ سپس در گام بعدی با استفاده از تابع ارزیاب هر کروموزوم در نسل نخست ارزیابی می‌شود. بر اساس مقادیر تابع ارزیاب، کروموزوم‌ها برای ایجاد نسل بعدی انتخاب می‌شوند. چند روش برای انتخاب کروموزوم‌های والدین وجود دارد. یکی از رایج‌ترین روش‌های انتخاب کروموزوم انتخاب چرخ رولت است که تابع ارزیاب کروموزوم را بر طبق بالاترین مقدار برای نسل‌های بعدی انتخاب می‌کند. نسل‌های بعدی براساس روش‌های الهام از تکامل طبیعی، مانند ترکیب و جهش تولید می‌شوند. در مرحلهٔ ترکیب، ویژگی‌های والدین برای ایجاد فرزندان ترکیب می‌شوند تا این‌که کروموزوم‌های بهتری ایجاد شوند. در ادامه در مرحلهٔ جهش در یک تعداد از مقادیر کروموزوم‌های مرحله ترکیب، تغییر ایجاد می‌شود تا از هم‌گرایی به بهینه محلی اجتناب شود و هم‌چنین تنوع در نسل بعدی ایجاد شود. امید است که جمعیت جدید نسبت به جمعیت قبلی بهتر باشد. این فرآیند تا برقراری شرطی که تعیین شده است، ادامه می‌یابد [16,15].

مجموعه‌داده‌های پزشکی دارای تعداد ویژگی‌های مختلفی است که سوابق بیمار را توصیف می‌کنند. برخی از این ویژگی‌ها ممکن است، در پیش‌بینی و تشخیص بیماری مؤثر نباشند یا تعدادی از ویژگی‌ها تأثیر مشابه با هم در پیش‌بینی و تشخیص بیماری داشته باشند. از طرفی مجموعه‌داده با تمام ویژگی‌های بیمار نیازمند حافظه و زمان بیشتر محاسباتی است. هم‌چنین ممکن است، دقّت طبقه‌بندی را کاهش دهند. ازاین‌رو تجزیه و تحلیل مجموعه‌داده با ابعاد بالا و تعداد نمونه کم در هوش مصنوعی اهمیت خاصی دارد. یکی از روش‌های مورد استفاده انتخاب ویژگی مؤثر با استفاده از الگوریتم ژنتیک است.

انتخاب ویژگی و تعیین عوامل مؤثر بیش‌تر به انتخاب یک زیرمجموعه از ویژگی‌های اولیه مربوط می‌شود که برای هدف مورد نظر اطلاعات لازم و کافی را دربر داشته باشد، به‌طوری‌که ویژگی‌های انتخاب‌شده معیار طبقه‌بندی را بهبود می‌دهند و پیش‌بینی و تشخیص بیماری با دقّت بالاتری انجام می‌شود. بنابراین انتخاب ویژگی و تعیین عوامل مؤثر کمک می‌کند تا ویژگی‌های افزونه و نامربوط کاهش یابند و هم‌چنین کمک به بهبود عملکرد طبقه‌بند می‌کند [17].

۳-۳- طبقه‌بند نزدیک‌ترین همسایه

الگوریتم نزدیک‌ترین همسایه برای یک داده آزمایشی به‌دنبال K نمونه از نزدیک‌ترین نمونه‌ها می‌شود. نزدیکی دو نمونه با به‌دست‌آوردن فاصلهٔ میان این دو نمونه محاسبه می‌شود. پس از یافتن این k دادهٔ مشابه با نمونهٔ آزمایشی، با رأی اکثریت برچسب طبقهٔ داده آزمایشی انتخاب می‌شود. چنان‌چه مقدار یک برای k تنظیم شود، در این‌صورت طبقهٔ نزدیک‌ترین داده به نمونه آزمایشی به‌عنوان طبقهٔ تخمینی ارائه می‌شود [18]. در پیش‌بینی بقای پنج‌سالهٔ کلیهٔ پیوندی، الگوریتم طبقه‌بند نخستین نزدیک‌ترین همسایه با معیار فاصلهٔ اقلیدسی به‌کار رفته است. مهم‌ترین معیار برای تعیین کارایی یک الگوریتم طبقه‌بندی، دقّت یا نرخ طبقه‌بندی است که این معیار، دقّت کل یک طبقه‌بند را محاسبه می‌کند. معیار دقّت براساس رابطهٔ (۱) محاسبه می‌شود:

$$percision = \frac{TP}{TP+FP} \qquad (۱)$$

علاوه‌بر معیار دقّت، معیار صحت طبقه‌بندی براساس رابطهٔ (۲) برای مقایسه روش پیشنهادی با کارهای پیشین لحاظ شده است.

$$accuracy = \frac{TP+TN}{TP+T\ \ +FP+FN} \qquad (۲)$$

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.29252/jsdp.15.4.85]

TN: بیان‌گر تعداد رکوردهایی است که دسته واقعی آنها منفی بوده و الگوریتم دسته‌بندی نیز دسته آنها را به‌درستی منفی تشخیص داده است.

TP: بیان‌گر تعداد رکوردهایی است که دستهٔ واقعی آنها مثبت بوده و الگوریتم دسته‌بندی نیز دستهٔ آنها را به‌درستی مثبت تشخیص داده است.

FP: بیان‌گر تعداد رکوردهایی است که دستهٔ واقعی آنها منفی بوده و الگوریتم دسته‌بندی، دستهٔ آنها را به‌اشتباه مثبت تشخیص داده است.

FN: بیان‌گر تعداد رکوردهایی است که دستهٔ واقعی آنها مثبت بوده و الگوریتم دسته‌بندی، دستهٔ آنها را به‌اشتباه منفی تشخیص داده است.

در این مطالعه معیار ارزیاب، دقّت طبقه‌بندی الگوریتم نخستین نزدیک‌ترین همسایه با اعتبارسنجی ضرب‌دری یکی بیرون به‌عنوان تابع شایستگی هر کروموزوم در نظر گرفته شده است. اعتبارسنجی ضرب‌دری یک روش ارزیابی است که مشخص می‌کند نتایج یک مدل تا چه اندازه قابل تعمیم و مستقل از داده‌های آموزشی است.

۴- الگوریتم پیشنهادی

در این مطالعه پیش‌بینی بقای پنج‌سالهٔ کلیهٔ پیوندی با استفاده از الگوریتم ژنتیک و الگوریتم نخستین نزدیک‌ترین همسایه انجام شده است. فرآیند کلی الگوریتم پیشنهادی در سه مرحله انجام شده است:

۱. جمع‌آوری داده
۲. پیش‌پردازش داده
۲،۱. نرمال‌سازی
۲،۲. متوازن‌سازی(روش بیش‌نمونه‌برداری و زیرنمونه‌برداری)
۳. مدل‌سازی(الگوریتم ژنتیک و الگوریتم اولین نزدیک‌ترین همسایه)

در ادامه هر مرحله به‌طورکامل شرح داده می‌شود.

۴-۱- جمع‌آوری داده

در این مرحله اطلاعات ۷۵۶بیمار کلیهٔ پیوندی با ۲۱ ویژگی از دو بیمارستان امام رضا(ع) و چهارمین شهید محراب کرمانشاه در سال‌های ۲۰۰۱-۲۰۱۲ جمع‌آوری شده است.

جدول (۱) ویژگی‌های بیماران کلیهٔ پیوندی را نشان می‌دهد. جنسیت دهنده، جنسیت گیرنده، نسبت فامیلی، همبستگی گروه خونی و نتیجه بقای پنج‌سالهٔ پیوند با مقادیر صفر و یک مقداردهی شده و مقادیر دیگر ویژگی‌ها مقادیر پیوسته حقیقی است که در مرحلهٔ پیش‌پردازش هنجارسازی شده است.

(جدول-۱): توصیف ویژگی‌های بیماران کلیهٔ پیوندی
(Table-1): description features of transplanted kidney patients

ردیف	نام ویژگی	ردیف	نام ویژگی
۱	وزن	۱۲	جنسیتِ دهنده
۲	سنّ ِدهنده	۱۳	جنسیتِ گیرنده
۳	سنّ ِگیرنده	۱۴	نسبت فامیلی
۴	اورهٔ قبل	۱۵	RH دهنده
۵	اورهٔ بعد	۱۶	RH گیرنده
۶	کراتین بعد پیوند	۱۷	سنّ دهندهٔ بالای ۳۰
۷	کراتین قبل پیوند	۱۸	بیماری اولیه
۸	هموگلوبین بعد پیوند	۱۹	سنّ گیرندهٔ بالای ۴۰
۹	هموگلوبین قبل پیوند	۲۰	کراتین بعد از پیوند
۱۰	هموتین بعد پیوند	۲۱	نتیجهٔ بقای پنج ساله پیوند
۱۱	هموتین قبل پیوند		

۴-۲- پیش‌پردازش داده

مرحلهٔ دوم، پیش‌پردازش داده است که یکی از بحرانی‌ترین مراحل موجود در فرآیند یادگیری ماشین و داده‌کاوی است. این مرحله شامل دو مرحلهٔ هنجارسازی و متوازن‌سازی داده‌های جمع‌آوری‌شدهٔ بیماران کلیهٔ پیوندی است.

۴-۲-۱- نرمال‌سازی

تولید نتایج کارا به داده‌های مناسب در مدل وابسته است. پیش‌پردازش داده‌ها با پاک‌سازی داده‌ها همراه است که برای کامل‌کردن داده‌های ناقص از میانگین و مد استفاده شده است. بعضی از مقادیر ویژگی‌های داده در محدوده یا دامنهٔ متفاوتی قرار دارند؛ ازاین‌رو برای یک‌پارچه‌سازی داده‌ها از رابطهٔ نرمال‌سازی زیر استفاده شده است.

$$Ni = (xi - \mu i)/\sigma i \qquad (۳)$$

که xi مقدارداده اولیه به‌ازای هر رکورد ، μi و σiبه‌ترتیب میانگین و انحراف معیارمقادیر هر ویژگی و Ni مقادیر نرمال شده است.

۴-۲-۲- متوازن‌سازی

مجموعه‌دادهٔ اولیهٔ جمع‌آوری‌شده از بیماران کلیهٔ پیوندی نامتوازن هستند. به‌طوری‌که از ۷۵۶ دادهٔ بیماران کلیهٔ پیوندی داده‌های متعلق به طبقهٔ قبول پیوند-طبقهٔکمینه-۹۲ رکورد و داده‌های متعلق به طبقهٔ رد پیوند-طبقهٔ بیشینه-۶۶۴ رکورد

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.29252/jsdp.15.4.85]

است. دراین‌صورت نرخ عدم توازن که برابر نسبت تعداد نمونه‌های طبقهٔ کمینه به تعداد نمونه‌های طبقهٔ بیشینه است در مجموعه‌دادهٔ مورد مطالعه دارای نرخ عدم توازن ۱۴درصد است.

با توجه به این‌که داده‌های جمع‌آوری‌شده نامتوازن هستند و عدم توازن سبب می‌شود هم‌گرایی مدل طبقه‌بندی به طرف طبقهٔ بیشینه صورت گیرد، که خود باعث کاهش کارایی طبقه‌بند می‌شود، لذا در مرحله پیش‌پردازشی برای متوازن‌سازی داده‌ها از روش‌های نمونه‌برداری استفاده شده است که دو روش بیش‌نمونه‌برداری و زیرنمونه‌برداری را شامل می‌شود.

۳-۴- مدل‌سازی

برای شناسایی عوامل مؤثر بقای پنج‌سالهٔ پیوند، مدلی براساس پیش‌بینی میزان بقای پنج‌سالهٔ پیوندکلیه ارائه شده است که دست‌یابی به نتایج تحلیلی مشابه با اصل و تمام داده‌ها تضمین می‌شود. از طرفی پیداکردن بهترین عوامل مؤثر از میان تمام زیر مجموعه‌های ممکن، کاری بغرنج است؛ لذا استفاده از الگوریتم‌های مناسب برای جستجو در فضای ویژگی‌ها و انتخاب عوامل مؤثر امری ضروری است. بنابراین در مدل پیشنهادی، الگوریتم ژنتیک برای جستجو در فضای ویژگی‌ها به‌کار رفته است.

در الگوریتم ژنتیک باید ابتدا مقادیر پارامترهای الگوریتم مقداردهی شوند. مقداردهی پارامترها شامل تعیین تعداد اعضای جمعیت، نرخ ترکیب، نرخ جهش و تعداد تکرارهای الگوریتم است که در این مطالعه به‌ترتیب از مقادیر ۲۰، ۰/۹۵، ۰/۰۱، ۳۰ استفاده شده است.

در این مدل به‌منظور شناسایی عوامل مؤثر برای بازنمایی هر کروموزوم از رشته‌ای از صفر و یک‌ها استفاده شده است که یک به معنی متغیر مؤثر در پیش‌بینی بقای پنج‌سالهٔ کلیهٔ پیوندی و صفر به معنی متغیر غیر مؤثر است. برای مثال رشته (۱۱۱۰۱۱۱۱۱۱۱۰۱۱۱۱۱۱۱۰) به‌عنوان یک کروموزوم در الگوریتم ژنتیک است که تمامی ویژگی‌ها به‌جز ویژگی‌های ۴،۱۲ و ۲۰ انتخاب شده‌اند. با تولید جمعیت اولیه(کروموزوم‌ها) به‌صورت رشته دودویی تصادفی، عوامل انتخاب‌شده توسط الگوریتم ژنتیک به طبقه‌بند نخستین نزدیک‌ترین همسایه داده می‌شود.

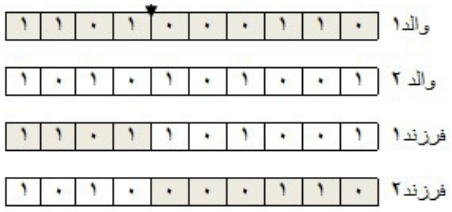
این طبقه‌بند با ویژگی‌های انتخابی موجود در هر کروموزوم، و اعتبارسنجی ضرب‌دری یکی بیرون میزان شایستگی هر کروموزوم را بر اساس دقّت پیش‌بینی بقای پنج‌سالهٔ کلیهٔ پیوندی تعیین می‌کند. به‌عبارتی مقدار تابع شایستگی(برازندگی)، میزان دقّت طبقه‌بند نخستین نزدیک‌ترین همسایه برای هر کروموزوم است که نشان‌دهندهٔ میزان مؤثربودن ویژگی‌های انتخابی در پیش‌بینی دادهٔ بیماران پیوند کلیه در طبقه‌بند نخستین نزدیک‌ترین همسایه است.

برای ساختن نسل بعد از عمل‌گرهای وراثتی، انتخاب، ترکیب و جهش استفاده می‌شود. ابتدا با عمل‌گر انتخاب چرخ رولت، کروموزوم‌ها با شایستگی بیشتر برای تولید مثل انتخاب می‌شوند؛ سپس عمل‌گر ترکیب دو کروموزوم والد را برای ساختن فرزندان ترکیب می‌کند تا کروموزوم‌های بهتری ایجاد شوند. در این مدل از تلفیق سه عمل‌گر ترکیب تک‌نقطه‌ای، دونقطه‌ای و یک‌نواخت استفاده شده است که در هر مرحله یکی از این سه عمل‌گر به‌صورت تصادفی برای ترکیب انتخاب می‌شود و دو فرزند برای نسل جدید تولید می‌کند [15]. در ترکیب تک‌نقطه‌ای یک نقطه به‌صورت تصادفی در کروموزوم‌های والد انتخاب‌شده و با یکدیگر ترکیب می شوند (شکل۱). در ترکیب دونقطه‌ای، دو نقطه به‌صورت تصادفی در کروموزوم‌های والد انتخاب شده و سپس ژن‌های بین نقاط انتخاب‌شده به‌صورت یک‌درمیان با یکدیگر عوض می‌شوند (شکل۲). در ترکیب یک‌نواخت، تولید فرزندان از روی ماسکی که به‌صورت تصادفی تولید می‌شود، انجام می‌گیرد. به این صورت که اگر در ماسک بیتی برابر یک باشد، والد نخست در تولید فرزند نخست و والد دوم در تولید فرزند دوم به‌کار می‌رود. اگر در ماسک بیتی برابر صفر باشد، والد نخست در تولید فرزند دوم و والد دوم در تولید فرزند نخست به‌کار می‌رود (شکل۳).

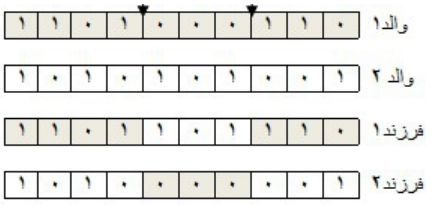
به‌کمک عمل‌گر جهش یک کروموزوم از نسل قبلی انتخاب و یک ژن از آن را به‌صورت تصادفی انتخاب و معکوس می‌کنیم. بعد از اعمال الگوریتم‌های ترکیب و جهش و با داشتن جمعیت قبلی، به‌دست‌آوردن کروموزوم‌های جدید و تعیین میزان شایستگی آن‌ها توسط طبقه‌بند نخستین نزدیک‌ترین همسایه کروموزوم‌ها با تابع شایستگی بیشتر، نسل بعد جمعیت را می‌سازند؛ این روال ادامه می‌یابد تا به تعداد مراحل مشخصی از تکرار الگوریتم برسد [21-19]؛ در انتها با رسیدن به آخرین تکرار و توقف الگوریتم ژنتیک ویژگی‌های مؤثر با بیشترین دقّت پیش‌بینی شناسایی می‌شوند.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

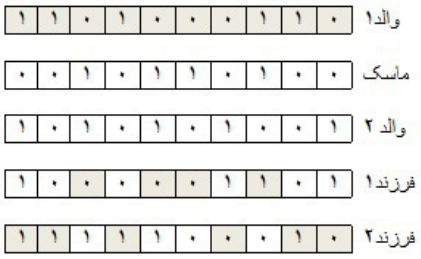
[DOI: 10.29252/jsdp.15.4.85]



(شکل-۱): ترکیب تک‌نقطه‌ای
(Figure-1): One-point crossover



(شکل-۲): ترکیب دو نقطه‌ای
(Figure-2): Two-point crossover



(شکل-۳): ترکیب یک-نواخت
(Figure-3): Uniform crossover

۵- یافته‌ها و بحث

مدل پیشنهادی در نرم‌افزار MATLAB روی لپ‌تاپی با پردازنده اینتل هفت‌هسته‌ای پیاده‌سازی شده است. مجموعه‌داده‌های جمع‌آوری‌شده که در بخش ۴-۱ بحث شد، برای ارزیابی عملکرد مدل پیشنهادی مورد استفاده قرار گرفت.

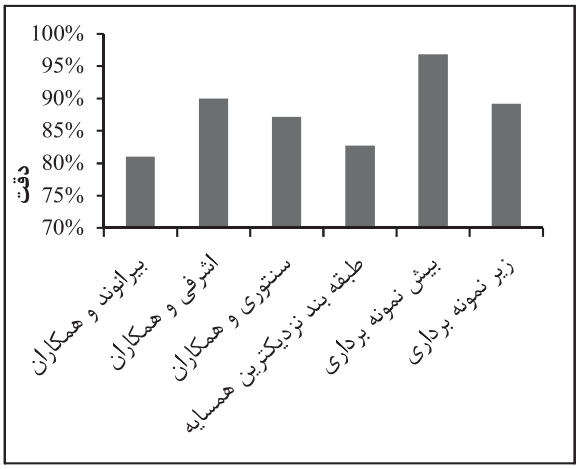
با توجه به این‌که فضای جستجوی داده‌ها نمایی است، لذا استفاده از الگوریتم جستجوی ژنتیک برای جستجوی چنین فضایی به‌منظور تعیین عوامل اثرگذار بر میزان بقای پنج‌سالهٔ کلیهٔ پیوندی در زمان معقول نقش حیاتی دارد. براین اساس با جستجوی فضای زیرمجموعه‌ها در داده‌های کلیهٔ پیوندی، مهم‌ترین عوامل مؤثر بر بقای پنج‌سالهٔ کلیه در روش بیش‌نمونه‌برداری و زیرنمونه‌برداری شناسایی شدند.

در مدل به‌دست‌آمده بعد از رسیدن الگوریتم ژنتیک به تعداد تکرار معین عوامل وزن، سنّ دهنده، سنّ گیرنده، اورهٔ قبل پیوند، کراتین قبل پیوند، هموگلوبین بعد پیوند، هموگلوبین قبل پیوند، جنسیت دهنده، RH دهنده و گیرنده، سنّ دهندهٔ بالای سی، بیماری اولیه، سنّ گیرندهٔ بالای چهل به‌عنوان ویژگی های تأثیرگذار در بقای پنج‌سالهٔ‌کلیهٔ پیوندی با دقّت ۹۶/۸ درصد در روش بیش‌نمونه‌برداری و عوامل کراتین قبل پیوند، هموگلوبین بعد پیوند، هموگلوبین قبل پیوند، هموتین بعد پیوند، هموتین قبل پیوند، کراتین بعد از پیوند با دقّت ۸۹/۲ درصد در روش زیرنمونه‌برداری شناسایی شدند؛ که در مقایسه با سایر روش‌های کارشده در حوزهٔ برآورد میزان بقای کلیهٔ پیوندی، از دقّت پیش‌بینی بالایی برخوردارند که می‌توان با درنظرگرفتن این عوامل به افزایش میزان بقای پنج‌سالهٔ کلیهٔ پیوندی کمک کرد.

سنتوری و همکاران در سال ۲۰۰۷ با استفاده از مدل یادگیری شبکهٔ عصبی برای پیش‌بینی نتیجه بقای کلیهٔ پیوندی اطفال را با دقّت ۸۷،۱۴ درصد پیش بینی کردند [6]. در سال ۲۰۰۹، اشرفی و همکاران با استفاده از شبکهٔ عصبی مدلی با دقّت نوددرصد جهت بررسی بقای کلیهٔ پیوندی در اصفهان ارائه دادند. هم‌چنین بر اساس نتایج این مطالعه نشان دادند، میزان بقای کلیهٔ پیوندی در پیوندهای زنده بیشتر از پیوندهای جسدی است [3]. در سال ۲۰۱۳ بیرانوند و همکاران با شبکهٔ عصبی پرسپترون چندلایه بر روی ۷۵۶ داده بیمارستان امام رضا(ع) و چهارمین شهید محراب کرمانشاه در سال‌های ۲۰۰۲-۲۰۱۲ بررسی کردند و به مدلی با دقّت ۸۱ درصد دست یافتند [9].

شکل (۴) مقایسه‌ای از کارهای اخیر با روش پیشنهادی در این مقاله (در دو حالت بیش‌نمونه‌برداری و زیرنمونه‌برداری داده‌ها) را نشان می‌دهد.

نتایج مقایسه حاکی از آن است که روش نمونه‌برداری در مقایسه با مطالعات پیشین، از نقطه‌نظر دقّت برتر است. با توجه به این‌که داده‌های بیرانوند و همکارانش با داده‌های استفاده‌شده در این مطالعه یکسان است، الگوریتم پیشنهادی در این مقاله با متوازن‌سازی داده‌ها، دقّت قابل توجه‌ای را برآورد می‌کند و با اطمینان خاطر بیشتری می‌توان از مدل پیشنهادی برای پیش‌بینی بقای کلیهٔ پیوندی استفاده کرد.



(شکل-۴): مقایسه روش‌های پیشنهادی با سایر روش‌ها
(Figure-4): Comparison of proposed methods with other methods

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.29252/jsdp.15.4.85]

میرزائی و همکاران درسال ۲۰۱۶، به‌کمک الگوریتم ژنتیک و روش هم‌جوشی اطلاعات متغیرهای تأثیرگذار را شناسایی کرده‌اند که این مدل دارای صحت پیش‌بینی ۹۱,۶۷ درصد است. در الگوریتم پیشنهادی با شناسایی متغیرهای تأثیرگذار توسط الگوریتم ژنتیک و طبقه‌بند نزدیک‌ترین همسایه صحت پیش‌بینی در روش بیش‌نمونه‌برداری و زیرنمونه‌برداری به‌ترتیب ۹۱/۸۹و ۸۳/۸۵درصد به‌دست آمده است [1]. نتایج مقایسه نشان می‌دهد که روش پیشنهادی با متوازن‌سازی داده‌ها براساس بیش‌نمونه‌برداری، سبب ایجاد مدل پیش‌بینی با صحت بالاتر شده است.

۶- نتیجه‌گیری

امروزه روش‌های آماری با افزایش تعداد مشاهدات و تعداد ویژگی‌ها یا عوامل مربوط به یک مشاهده کارایی خود را از دست داده‌اند. ابزارهای یادگیری ماشین و داده‌کاوی کوششی برای به‌دست‌آوردن اطلاعات مفید از میان این داده‌هاست.

این مطالعه با رویکرد شناسایی عوامل مؤثر و پیش‌بینی بقای پنج‌سالهٔ کلیهٔ پیوندی در داده‌های نامتوازن به‌کمک الگوریتم ژنتیک و روش یادگیری طبقه‌بند نزدیک‌ترین همسایه انجام شده است. با توجه به این‌که داده‌های آموزشی توزیع طبقهٔ نامتوازنی دارند و طبقه‌بند یادگرفته‌شده اغلب به طبقهٔ بیشینه متمایل می‌شود، این موضوع به پیش‌بینی بسیار ضعیفی از طبقهٔ کمینه منجر می‌شود؛ لذا برای بهبود کارایی از دو روش متوازن‌سازی بیش‌نمونه‌برداری و زیرنمونه‌برداری استفاده شد. از طرف دیگر افزایش تعداد ویژگی‌ها و عوامل، چالش‌های محاسباتی بالایی را به‌دنبال دارد و منجر به یافتن الگوهای نامعتبر می‌شود. بنابراین انتخاب عوامل مؤثر نقش مهمی را در تعیین عملکرد پیش‌بینی الگوریتم نزدیک‌ترین همسایه به‌عنوان یک مدل یادگیری بازی می‌کنند. با انتخاب عوامل مؤثر و کاهش ابعاد، داده‌ها علاوه‌بر کاهش زمان محاسبات، دقّت پیش‌بینی بر میزان بقای پنج‌سالهٔ کلیهٔ پیوندی افزایش یافته است.

همان‌طور که مشاهده شد در هر دو روش بیش‌نمونه‌برداری و زیرنمونه‌برداری دقّت پیش‌بینی میزان بقای پنج‌سالهٔ کلیهٔ پیوندی افزایش یافته است که نشان‌دهندهٔ بهینه‌بودن روش پیشنهادی است. علاوه‌بر آن، روش بیش‌نمونه‌برداری نسبت به روش زیرنمونه‌برداری دقّت و صحت بهتری را داشته و روش مناسبی در جهت شناسایی عوامل مؤثر در پیش بینی بقای پنج‌سالهٔ کلیهٔ پیوندی است.

۷- مراجع

7- References

[۱] میرزایی، محترم، و فیروز آبادی، سید محمد، "کاربرد داده‌کاوی در پیش‌بینی بقای پیوند کلیه و شناسایی متغیرهای تأثیرگذار در بقای کلیهٔ پیوندی"، مجله انفورماتیک سلامت و زیست پزشکی، مرکز تحقیقات انفورماتیک پزشکی، دوره ۳، شماره ۱، صفحات ۱-۹، ۱۳۹۵.

[1] M. Mirzaei, and M. Firooz Abadi, "The impact of data mining on prediction of renal transplantation survival and identifying the effective factors on the transplanted kidney," *Journal Of Health and Biomedical Informatics, Medical Informatics Research Center*, vol. 3, pp. 1-9, 2016.

[۲] جوانروح گیوی، نیلوفر، علیمی، رسول، اسماعیلی، حبیب ا...، شاکری، محمدتقی، و شمسا، علی، "عوامل مؤثر بر بقای پیوند کلیه و برآورد خطر رد پیوند برای پیوند شدگان مراجعه کننده به بیمارستان قائم مشهد"، مجله دانشگاه علوم پزشکی خراسان شمالی، دوره ۵، شماره ۲، صفحات ۳۱۵-۳۲۱، ۱۳۹۲.

[2] N. Javanroh Givi, R. Alimi, H. Esmaily, M. T. Shakeri, and A. Shamsa, "Assessment of effective factors on renal transplantation estimation of rejection hazard for transplanted in Mashhad Qaem hospital," *Journal of North Khorasan University of Medical Sciences*, vol. 5, pp. 315-321, 2013.

[۳] اشرفی، مهدی، و همکاران،" پیش‌بینی بقای پنج ساله پیوند کلیه با استفاده از مدل شبکهٔ عصبی مصنوعی: گزارش ۲۲ سال پی‌گیری از ۳۱۶ بیمار در اصفهان"، مجله دانشکده پزشکی، دانشگاه علوم پزشکی تهران، دوره ۶۷، شماره ۵، صفحات ۳۵۳-۳۵۹، ۱۳۸۸.

[3] M. Ashrafi, and et. al, "Application of artificial neural network to predict graft survival after kidney transplantation: reports of 22 years follow up of 316 patients in Isfahan," *Tehran University Medical Journal*, vol. 67, pp. 353-359, 2009.

[۴] الماسی حشیانی، امیر، رجایی‌فرد، عبدالرضا، حسن‌زاده، جعفر، وصلاحی، حشمتا...،" تحلیل بقا پیوند کلیه و ارتباط آن با سن و جنس دهنده و گیرنده عضو بین بیماران پیوند شده"، مجله علمی دانشگاه علوم پزشکی سمنان، جلد ۱۱، شماره ۴، صفحات ۳۰۲-۳۰۷، ۱۳۸۹.

[4] A. Almasi Hashiani, A. Rajaeefard, J. Hassanzade, and H. Salahi, "Survival analysis of renal transplantation and its relationship with age and sex," *Koomesh*, vol. 11, pp. 302-307, 2010.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.29252/jsdp.15.4.85]

[5] T. D. Noia, and et al, "An end stage kidney disease predic-tor based on an artificial neural networks ensemble," *Expert Systems with App-lications*, vol. 40, pp. 4438-4445, 2013.

[6] G. Santori, I. Fontana, and U. Valente, "Application of an artificial neural network model to predict delayed decrease of serum creatinine in pediatric patients after kidney transplantation," *Transplant Proc*, vol. 39, pp. 1813-1819, 2007.

[7] T. S. Brown, and et al, "Bayesian modeling of pre transplant variables accurately predicts kidney graft survival," *Am J Nephrol*, vol. 6, pp. 561-569, 2012.

[8] J. Lasserre, S. Arnold, M. Vingron, P. Reinke, and C. P. Hinrichs, "Predicting the outcome of renal transplantation," *J Am Med Inform Assoc*, vol. 19, pp. 255-262, 2012.

[9] A. H. Hashemian, B. beiranvand, M. Rezaei, A. Bardideh, and E. Zand-Karimi, "Comparison of artificial neural network of kidney transplant survival*," International Journal of Advanced Biological and biomedical Research*, vol. 1, pp. 1204-1212, 2013.

[10] R. J. Oskouei and B. S. Bigham, "Over-sampling via under-sampling in strongly Imbalanced data," International Journal of Advanced Intelligence Paradigms, 2015.

[11] M. M. Rahman, and D. N Davis, "Addressing the class imbalance problem in medical datasets," *Int J Machine Learning and Compute*, vol. 2, pp. 224-228, 2013.

[12] N. V Chawla, "Data mining for imbalanced datasets: an overview," *Data Mining Knowledge Discovery Handbook*, 2005.

[13] Y. Sun, A. K. C. Wong, and M. S Kamel, "Classification of imbalanced data: a review," *Int j patt Recogn Artif Intell*, vol. 4, pp. 687-719, 2009.

[14] D. C. Li, C. W. Liu, and S. C. Hu, "A learning method for the class imbalance problem with medical datasets," *J comput Bio Medi*, vol. 5, pp. 509-518, 2010.

[۱۵] حسین‌خانی، فاطمه و ناصرشریف، بابک،" دو روش تبدیل ویژگی مبتنی بر الگوریتم‌های ژنتیک برای کاهش خطای دسته‌بندی ماشین بردار پشتیبان"، مجله پردازش علائم و داده‌ها، جلد ۲۴، شماره ۲، صفحات ۲۳-۳۹، ۱۳۹۴.

[15] F. Hoseinkhani, and B. Naser Sharif, " Two methods of converting feature based on genetic algorithms to reduce the classification error of support vector machine," *Journal Of signs and data Processing*, vol. 24, pp. 23-39, 2015.

[16] H. Hoglund, " Tax payment default prediction using genetic algorithm-based variable selection," *Expert Syst Appl*, vol. 88, pp. 368-375, 2017.

[17] S. Nagpal, S. Arora, S. Dey and Shreya, " Feature selection using gravitational search algorithm for biomedical data," *Procedia Comput Sci*, vol. 115, pp. 258-265, 2017.

[18] X. We and et. al, "Top 10 algorithms in data mining," *knowl Inf Syst*, vol. 14, pp. 1-37, 2008.

[19] J. H. Holland, " Adaptation in natural and artificial systems," *University of Michigan Press*, 1975.

[20] D. E. Goldberg, "Genetic algorithms in search optimization and mechine learning," *Addison-Wesley Publishing, INC. Reading. Mass*, 1989.

[21] S. Olariu and A. Y. Zomaya, "Handbook of bioinspired algorithms and applications," *Taylor & Francis Group, LLC Press*, 2006.

نسیبه امامی مدارک کارشناسی و کارشناسی ارشد خود را به ترتیب از دانشگاه یزد و شهید باهنر کرمان در رشته علوم کامپیوتر گرایش سامانه‌های هوشند اخذ کرده است. وی هم‌اکنون عضو هیأت علمی دانشگاه کوثر بجنورد است و در زمینه‌های یادگیری ماشین، هوش مصنوعی، بهینه‌سازی و داده‌کاوی فعالیت دارد.

نشانی رایانامهٔ ایشان عبارت است از:

nasibeh.emami@kub.ac.ir

زینب حسنی مدارک کارشناسی و کارشناسی ارشد خود را به‌ترتیب از دانشگاه یزد و مرکز تحصیلات تکمیلی زنجان در رشته علوم کامپیوتر گرایش سامانه‌های هوشند اخذ کرده است. وی هم‌اکنون عضو هیأت علمی دانشگاه کوثر بجنورد است و در زمینه‌های یادگیری ماشین، هوش مصنوعی، بهینه‌سازی و هندسه محاسباتی فعالیت دارد.

نشانی رایانامهٔ ایشان عبارت است از:

Hassani@kub.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.29252/jsdp.15.4.85]

[DOI: 10.29252/jsdp.15.4.85] [Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]