پیش بینی آسیب *DNA* و کیفیت جنین در لقاح مصنوعی بر مبنای داده های مرکز رویان با استفاده از تکنیک های هوش مصنوعی

لاله آرمی[۱]، روشنک رضائی*[۲] ، محمدنوید بقولی زاده[۳]

[۱] دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران، laleharmi93@gmail.com

*[۲]مدیرعامل شرکت مهندسی داده پردازان اندیشه نوین آناهیتا ، rr28960@gmail.com

[۳]محمدنوید بقولی زاده، n.baghoulizadeh@gmail.com

چکیده

لقاح مصنوعی یک تکنیک پرکاربرد در پزشکی تولید مثل است، اما پیش‌بینی کیفیت جنین و شناسایی آسیب‌های احتمالی *DNA* چالشی در این فرآیند است. برای رفع این مشکل، محققان از الگوریتم‌های هوش مصنوعی استفاده کرده‌اند تا مدل‌هایی توسعه دهند که کیفیت جنین را پیش‌بینی و آسیب *DNA* را تشخیص دهند. با استفاده از تکنیک‌های یادگیری ماشین، این مدل‌ها می‌توانند الگوها و همبستگی‌هایی را جهت شناسایی آسیب دیدگی *DNA* و کیفیت جنین شناسایی کنند و پیش‌بینی‌هایی در مورد کیفیت جنین و آسیب *DNA* ایجاد کنند و به پزشکان اجازه دهند تدابیر مناسبی را برای کاهش آسیب *DNA* و افزایش کیفیت جنین اتخاذ کنند و با تصمیمات آگاهانه، شانس یک نتیجه موفقیت‌آمیز را بهینه کنند. در این چالش بر اساس داده‌های دریافت‌شده از مرکز رویان سعی در پیش‌بینی آسیب *DNA* براساس ارزیابی نمونه اسپرم، سن مردان و همچنین پیش‌بینی نتایج لقاح، کیفیت جنین و حاملگی زوجین با استفاده از نتایج آسیب *DNA* در قالب دو زیر مساله مطرح شده را داریم که با توجه به تکنیک‌ها و روش‌های در حوزه هوش مصنوعی مسائل این چالش را پوشش خواهیم داد.

کلید واژه‌ها: آسیب DNA، .آ حاملگی زوجین، دابوست، رگرسیون جنگل تصادفی

اعتبارسنجی الگوریتم ها بر روی داده‌های آزمایشگاهی و نتایج واقعی جمع آوری شده از بیماران می تواند به محققان کمک کند تا الگوهای مشخصی را برای تشخیص و پیش‌بینی آسیب DNA ان ای تعریف کنند. از این طریق می توانند به موقع اطلاع داشته باشند که کدام بیماران بیشترین خطر بروز آسیب دی ان ای را دارند و اقدامات مناسب جلوگیری از آسیب دی ان ای را انجام دهند.
علاوه بر آن، تحقیقات در زمینه پیش بینی کیفیت جنین نیز انجام می شود. با استفاده از داده های جمع آوری شده از بیماران و تکنیک‌های هوش مصنوعی، می‌توان الگوها و روش‌هایی را برای تشخیص جنین‌های با کیفیت بالا و کاهش احتمال جنین‌های ناسالم تعریف کرد. این الگوها و روش‌ها به پزشکان در تصمیم گیری در مورد انتخاب جنین بهتر و کاهش خطرات مربوط به لقاح مصنوعی کمک می کنند.
چندین مطالعه نتایج امیدوارکننده‌ای را در استفاده از هوش مصنوعی برای پیش‌بینی کیفیت جنین و آسیب DNA نشان داده‌اند. به عنوان مثال، یک مطالعه از الگوریتم های یادگیری عمیق برای پیش بینی کیفیت جنین بر اساس پارامترهای متعدد، مانند ویژگی‌های مورفولوژیکی و اطلاعات ژنتیکی استفاده کرد. این مدل به دقت بالایی در پیش بینی احتمال لانه گزینی جنین و موفقیت بارداری دست یافت.
مطالعه دیگری از هوش مصنوعی برای تشخیص آسیب DNA در نمونه های اسپرم استفاده کرد. با تجزیه و تحلیل تصاویر سلول های اسپرم، این مدل توانست میزان تکه تکه شدن DNA را که با رشد ضعیف جنین و کاهش باروری مرتبط است، تعیین کند. این اطلاعات می تواند به پزشکان کمک کند تا بهترین نمونه های اسپرم را برای لقاح مصنوعی انتخاب کنند. به طور کلی، تحقیقات در زمینه استفاده از هوش مصنوعی در پیش بینی آسیب دی ان ای و کیفیت جنین در لقاح مصنوعی به منظور بهبود نتایج این فرآیند و کاهش خطرات آن صورت می گیرد. این تحقیقات می‌توانند به محققان و پزشکان در انتخاب بهتر جنین و پیش بینی موارد ناسازگاری به آنها کمک کنند.

۳-پیش‌نیازها

در این بخش به طور مختصر به روش‌ها و تکنیک های یادگیری ماشین پرداخته شده تا زمینه‌های لازم برای روش پیشنهادی فراهم شود.

۳-۱-روش های برای تخمین داده های ناموجود

در بسیاری از مجموعه داده‌ها، داده‌هایی که برای یک یا چند ویژگی موجود نیستند یا به عبارت دیگر داده‌های گم‌شده، مشکلاتی را ایجاد می‌کنند. این داده‌های گم‌شده می‌توانند باعث کاهش دقت و کارایی مدل‌های یادگیری ماشین شوند و باعث تحلیل نادرست داده‌ها و انجام پیش‌بینی‌های ناصحیح شوند[1-3]. برای این منظور، تخمین داده‌های گم‌شده و پر کردن آن‌ها با مقادیر تخمین زده شده، بهبود دقت و کارایی مدل‌های یادگیری ماشین و تحلیل دقیق‌تر داده‌ها کمک می‌کند و همچنین جلوگیری از بروز خطاهای غیرمنتظره و افزایش دقت تحلیل‌های می‌شود. از جمله روش‌های که برای تخمین داده‌های ناموجود به کاربرده‌می‌شود میتوان به KNN Imputation

۱-مقدمه

لقاح مصنوعی یک تکنیک رایج در پزشکی تولید مثل برای تسهیل لقاح است. با این حال، یکی از چالش‌های تلقیح مصنوعی، پیش‌بینی کیفیت جنین و شناسایی آسیب‌های احتمالی DNA است که می‌تواند بر موفقیت این روش تأثیر زیادی بگذارد. یکی از مشکلات شایع در فرایند لقاح مصنوعی، آسیب دیدگی دی ان ای و کیفیت جنین است که می تواند تاثیری مهم بر موفقیت و نتیجه حاصل از لقاح داشته باشد. در حال حاضر، شناخت دقیقی از عوامل تاثیرگذار و میزان تاثیر آن‌ها بر آسیب دی ان ای و کیفیت جنین وجود ندارد که موجب کاهش اطمینان و قابلیت پیش بینی نتیجه لقاح می شود.
برای رفع این مشکل، محققان به الگوریتم‌های هوش مصنوعی[1] روی آورده‌اند تا مدل‌هایی را توسعه دهند که می‌توانند کیفیت جنین را به‌طور دقیق پیش‌بینی کنند و آسیب DNA را تشخیص دهند. این مدل‌ها از مجموعه داده‌های بزرگ از چرخه‌های لقاح مصنوعی قبلی، از جمله اطلاعات مربوط به بیماران، کیفیت اسپرم و تخمک و جنین‌های حاصل استفاده می‌کنند.
با استفاده از تکنیک‌های یادگیری ماشین، این مدل‌های هوش مصنوعی می‌توانند حجم وسیعی از داده های جمع آوری شده از فرایند لقاح مصنوعی را تجزیه و تحلیل کنند و الگوها و همبستگی‌هایی را جهت بررسی آسیب دیدگی دی ان ای و کیفیت جنین شناسایی کنند که تشخیص آن‌ها برای انسان دشوار است. سپس این مدل‌ها می‌توانند پیش‌بینی‌هایی در مورد کیفیت جنین و آسیب DNA ایجاد کنند و به پزشکان اجازه می‌دهند تدابیر مناسبی را برای کاهش آسیب دی ان ای و افزایش کیفیت جنین اتخاذ کنند و با تصمیمات آگاهانه، شانس یک نتیجه موفقیت‌آمیز را بهینه کنند.
به طور کلی، ادغام هوش مصنوعی در لقاح مصنوعی پتانسیل ایجاد انقلابی در زمینه پزشکی تولید مثل را دارد. با استفاده از قدرت یادگیری ماشینی، پزشکان می توانند میزان موفقیت روش های لقاح مصنوعی را بهبود بخشند و مراقبت های کارآمدتر و شخصی تر را به بیماران ارائه دهند. با این حال، تحقیقات و اعتبار سنجی بیشتر برای اطمینان از صحت و قابلیت اطمینان این مدل‌های هوش مصنوعی قبل از اجرای گسترده در عمل بالینی ضروری است.
ساختار این مقاله به شرح زیر است، در بخش دوم مروری بر ادبیات خواهیم داشت در بخش سوم روش های پایه و اولیه که برای حل دو زیر مسأله بوده است معرفی می‌گردد. در بخش چهارم به معرفی روش استفاده شده برای دو زیر مساله پرداخته و نتایج هر یک را در زیر بخش خودش نشان داده است و در بخش نهایی به بررسی و جمع بندی می پردازیم .

۲-مروری بر روی ادبیات تحقیق

تحقیقات در زمینه پیش بینی آسیب DNA و کیفیت جنین در لقاح مصنوعی با استفاده از هوش مصنوعی به منظور بهبود فرآیند لقاح مصنوعی و کاهش خطرات آن انجام شده است. این تحقیقات به دنبال استفاده از الگوریتم‌ها و مدل‌های هوش مصنوعی مانند شبکه های عصبی و الگوریتم ژنتیک برای پیش-بینی عملکرد لقاح مصنوعی و کیفیت جنین است.
یکی از موضوعات مهم در این تحقیقات، پیش بینی آسیب DNA است.

[1] Artificial intelligence (AI)

۳-۲- روش‌های مبتنی بر یادگیری مرکب

استفاده از یادگیری مرکب در حوزه یادگیری ماشین از طبیعت انسان برای تصمیم گیری‌های مهم الهام گرفته است، به این صورت که نظرات افراد مختلف پرسیده شده و برای نظرات آن‌ها، وزنی درنظر گرفته می‌شود. سپس نظرات با قاعده خاصی ترکیب شده تا تصمیم معقول اتخاذ شود. یادگیری مرکب رویکردی جهت بهبود عملکرد درطبقه بندی است[12-14].

۳-۲-۱-روش آدابوست

الگوریتم آدابوست، یک روش یادگیری مرکب است و معروف ترین الگوریتم از خانواده الگوریتم های تقویتی است که توسط یاو فروند و رابرت شاپیر ارائه شده است[15]. در روش آدابوست، طبقه‌بندی کننده هر مرحله جدید به نفع نمونه-های غلط طبقه‌بندی شده در مراحل قبل تنظیم می‌گردد. به عبارت دیگر، در هر تکرار، تمرکز طبقه بند مربوطه بیشتر بر تشخیص نمونه‌هایی است که به اشتباه توسط طبقه‌بندهای مراحل قبل برچسب گذاری شده اند. طبقه‌بندی پایه که در اینجا استفاده می‌شود فقط کافیست از طبقه‌بندی تصادفی (٪ ۵۰) بهتر باشد؛ و به این ترتیب بهبود عملکرد الگوریتم با تکرارهای بیشتر بهبود می‌یابد. روند کلی این الگوریتم به این صورت است که تعداد زیادی طبقه‌بندی پایه که طبقه بندی بهتری ارائه نموده اند را تشویق کرده و وزن بیشتری به آن طبقه‌بندی اختصاص می‌دهد. همچنین به طبقه‌بندی بدتر ارائه شده وزن کمتری اختصاص می دهد.

۳-۲-۲- رگرسیون جنگل تصادفی

RandomForestRegressor یک الگوریتم یادگیری ماشین است که برای مسائل پیش بینی و رگرسیون استفاده می شود. این الگوریتم از مجموعه ای از درخت های تصمیم استفاده می‌کند تا مدلی قوی و پایدار برای پیش بینی مقادیر عددی ایجاد کند[16-19]. روش RandomForestRegressor از تکنیک تصادفی جهت داده ها (Random Forest) استفاده می کند که با انتخاب تصادفی زیرمجموعه‌هایی از داده‌ها و ساخت درخت‌های تصمیم مستقل، مدلی قوی و انعطاف پذیر ایجاد می کند. سپس با میانگین گیری از پیش‌بینی-های تمام درخت‌ها، مقدار نهایی پیش بینی شده را تولید می‌کند. این الگوریتم از ویژگی های مختلف داده ها برای ساختن درخت های تصمیم استفاده می کند و به این ترتیب می تواند با داده های پرت و نویزی نیز مقاوم باشد. همچنین این الگوریتم از روش تقویت کننده یادگیری ماشین استفاده می کند که باعث می‌شود مدل نهایی بسیار قوی و دقیق باشد. این الگوریتم به دلیل اینکه از ترکیب چندین درخت تصمیم استفاده می‌کند، دارای انعطاف‌پذیری بالا و مقاومت در برابر بیش‌برازش[3] است. همچنین، این الگوریتم می‌تواند با داده‌های بزرگ و با ابعاد بالا نیز مقابله کند و به خوبی عملکرد کند.

۳-۲-۳- اختلاط خبره ها بر پایه ی ماشین یادگیری سریع با شبکه میانجی آموزش پذیر

در [20,21] روش اختلاط خبره مبتنی بر ELM با شبکه میانجی آموزش پذیر که به اختصارMEETG نامیده شده است. روش MEETG یک روش [4,5]، Denoise Autoencoder [6]، تجزیه ماتریس غیرمنفی [7,8]و غیره اشاره نمود[9-11] .

۱-پر کردن مقدار گم‌شده با استفاده از KNN:

KNN Imputation یا پر کردن مقدار گم‌شده با استفاده از k-nearest neighbors یک تکنیک است که برای پر کردن مقادیر گم‌شده در یک مجموعه داده با تخمین آن‌ها بر اساس ویژگی‌های داده‌های نزدیک مشابه استفاده می‌شود. این روش به ویژه برای مجموعه داده‌هایی با تعداد کمی مقادیر گم‌شده مؤثر است و زمانی کارایی دارد که مقادیر گم‌شده به صورت تصادفی پخش شده باشند. این روش با پیدا کردن k نزدیک‌ترین نقاط داده (همسایه‌های نزدیک) به مشاهده با مقدار گم‌شده کار می‌کند و سپس با استفاده از مقادیر مشخص شده آن‌ها، مقدار گم‌شده را تخمین می‌زند. پس از شناسایی همسایه‌های نزدیک، با استفاده از مقادیر شناخته شده آن‌ها برای ویژگی گم‌شده، مقدار ناموجود تخمین زده می‌شود. برای تخمین مقدار گم‌شده می‌توان از میانگین، مد، یا میانگین وزنی مقادیر شناخته شده از همسایه‌های نزدیک استفاده کرد.

۲-دی‌نویز اتوانکودر :

Denoise Autoencoder یک تکنیک یادگیری عمیق است که برای پر کردن مقادیر گم‌شده با آموزش یک شبکه عصبی برای بازسازی داده ورودی با مقادیر گم‌شده استفاده می‌شود. این روش با یادگیری الگوها و ارتباطات پایه‌ای در داده‌ها برای پر کردن مقادیر گم‌شده کارایی دارد. این روش برای مجموعه داده‌هایی با تعداد زیادی مقادیر گم‌شده مؤثر است و می‌تواند الگوهای پیچیده در داده‌ها را مدیریت کند. اتوانکودر برای کمینه کردن تفاوت بین داده ورودی و داده بازسازی شده آموزش داده می‌شود و به این ترتیب می‌آموزد که با توجه به الگوهای یادگرفته شده، مقادیر گم‌شده را پر کند.

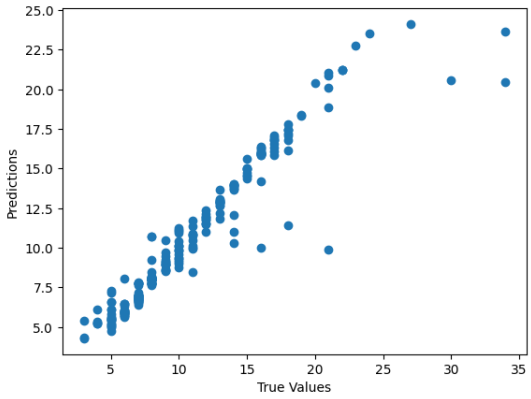
۳ تجزیه ماتریس غیرمنفی:

تجزیه ماتریس، به ویژه تجزیه ماتریس غیرمنفی[2]، یک تکنیک کاهش ابعاد است که همچنین برای برآورد مقادیر گم‌شده استفاده می‌شود. این روش با تجزیه ماتریس اصلی داده به دو ماتریس کم‌ابعاد که ویژگی‌های پنهان داده را ضبط می‌کنند کار می‌کند. این روش برای مجموعه داده‌هایی با تعداد زیادی مقادیر گم‌شده مؤثر است و می‌تواند روابط غیرخطی در داده‌ها را مدیریت کند. هدف NMF پیدا کردن دو ماتریس غیر منفی است که حاصل ضرب آن‌ها به داده اصلی نزدیک باشد و به این ترتیب ساختار پنهان داده را در داده‌ها ضبط کرده و مقادیر گم‌شده را بر اساس این ساختار پنهان پر می‌کند.

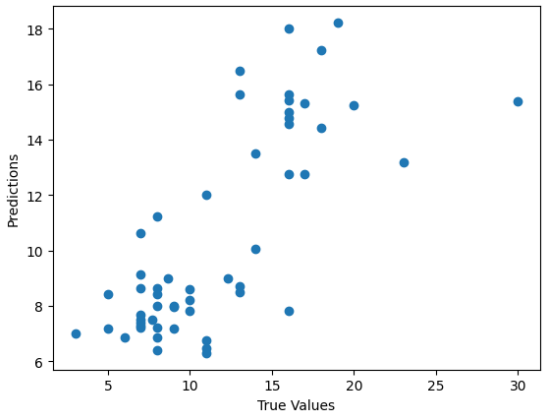
لازم به ذکر است که روش KNN Imputation مناسب برای مجموعه داده‌هایی با تعداد کمی مقادیر گم‌شده است و زمانی کارایی دارد که مقادیر گم‌شده به صورت تصادفی پخش شده باشند. دی‌نویز اتوانکودر برای مجموعه داده‌هایی با تعداد زیادی مقادیر گم‌شده مؤثر است و می‌تواند الگوهای پیچیده در داده‌ها را مدیریت کند. تجزیه ماتریس NMF برای مجموعه داده‌هایی با تعداد زیادی مقادیر گم‌شده مؤثر است و می‌تواند روابط غیرخطی در داده‌ها را مدیریت کند. هر یک از این متدها دارای قدرت‌ها و مزایا مختلفی هستند و می‌توانند براساس ویژگی‌ها و نیازهای خاص مجموعه داده انتخاب شوند.

[2] Negative Matrix Factorization (NMF)

[3] overfitting



شکل ۱: ماتریس درهم‌ریختگی برای پیش‌بینی حاملگی مثبت و منفی



۴-۲- پیش‌بینی پیش بینی نتایج لقاح، کیفیت جنین، و حاملگی زوجین با استفاده از نتایج اسیب DNA

در این چالش ابتدا با استفاده از KNN Imputation برای تخمین داده‌های ناموجود استفاده نموده‌ایم. سپس بعد از تقسیم نمودن به دادهای آموزشی و تست برای پیش‌بینی سه ویژگی Number.of.emabryo.with.gradA، of.emabryo.with.gradB و of.emabryo.with.gradC از روش RandomForestRegressor استفاده نموده و از تقسیم مقادیر بدست-آمده‌به Total.number.of.embryos برای ویژگی‌های Percent.of.emabryo.with.gradA ، with.gradB و with.gradC انجام داده‌ایم و برای پیش‌بینی حاملگی مثبت و عدد دو نشانگر حاملگی منفی از طبقه بندهای مرکب همچون آدابوست و MEETG استفاده شده است که خروجی‌های هر دو مدل یکسان بوده است.

روش فوق برای دیتاست آمورش که در اختیار قرار گرفته‌شده که شامل ۷۰ نمونه بوده است را ۲۱ نمونه را برای تست اختصاص داده‌ایم. در ادامه نتایج یادگیری مرکب است که بر اساس ترکیبی از خبره‌ها برای بهبود هزینه محاسبه و تسریع فرایند یادگیری ME عمل می‌نماید. یادگیری مرکب رویکردی است که توانایی بالایی در بهبود دقت طبقه‌بندی کلی دارد. روشMEETG از مزایای ME و ELM استفاده می‌کند درمعماری این روش متشکل از چندین متخصص و یک شبکه میانجی است فضای ورودی مسئله تجزیه می شود و کارشناسان با مدیریت شبکه میانجی به فضاهای فرعی مختلف اختصاص می یابند. این روش در یک دور آموزش می‌بیندو نیاز به آموزش مکرر و تکراری نیست بنابراین فرآیند یادگیری بسیار سریع‌تر و توانایی تعمیم بهتر را در مقایسه با سایر الگوریتم‌های یادگیری سنتی ارائه می‌کند [22,23].

۴-روش پیشنهادی

برای دو مساله های مطرح شده با توجه به دیتاست دریافتی (داده های مرکز رویان) که این داده ها برخی ناموجود بودن که لازم است یک پیش‌پردازشی روی داده ها انجام شد. ما در این چالش با توجه به مبحث بخش سوم اشاره کردیم از روش KNN Imputation برای تخمین داده‌های ناموجود استفاده نموده‌ایم. این روش مناسب برای مجموعه داده‌هایی با تعداد کمی مقادیر گم‌شده است و زمانی کارایی دارد که مقادیر گم‌شده به صورت تصادفی پخش شده باشند.

۴-۱- پیش‌بینی آسیب DNA با استفاده از ارزیابی نمونه اسپرم و سن مردان

دراین مساله هدف پیش‌بینی ویژگی‌هایSCSAassay،satiability.HDS وTUNELassay که نشان دهنده آسیب DNA هستند می‌باشد. برای این قسمت از چالش از روش RandomForestRegressor برای پیش‌بینی آسیب DNA استفاده نمودیم و معیار MSE را مدل ارزیابی خود انتخاب نمودیم.

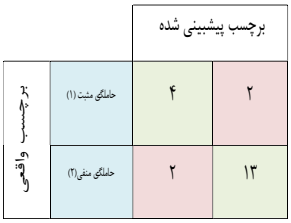
MSE به معنای میانگین مربعات خطا[4] است و یک معیار اندازه‌گیری خطا در مدل‌های پیش‌بینی و تخمین است. این معیار برای اندازه‌گیری فاصله بین مقادیر پیش‌بینی شده توسط یک مدل و مقادیر واقعی داده‌ها استفاده می‌شود. MSE برابر با جمع مربعات اختلاف بین مقادیر پیش‌بینی شده و مقادیر واقعی داده‌ها است و با کاهش این مقدار، دقت و عملکرد مدل بهبود می‌یابد.

در پارت اول دیتاستی با ۱۷۶ نمونه و ۱۳ ویژگی و ۳ ویژگی به عنوان هدف دریافت شده است که برای آموزش مدل استفاده شده است. ما برای آموزش و همچنین تنظیم پارامترهای مدل خود از این دیتا دریافت شده آن را به ۹۰٪ نمونه های آموزش، و ۱۰٪ دیگر را برای نمونه‌های ارزیابی تقسیم نمودیم.

خروجی نتایج در مرحله اول با میانگین مربع خطا برابر با که۱۲.۷ که نمودار خروجی‌های تست در مرحله اول در شکل ۱ نمایش داده‌شده‌است. در ادامه، بعد از در اختیار گرفتن داده های تست که شامل نمونه ۷۶ می‌باشد را روی مدل ذخیره شده خود اجرا گرفتیم.

[4] Mean squared error

حاصل که به صورت ماتریس در هم‌ریختگی در شکل ۳ نمایش داده‌شده‌است. با استفاده از ماتریس درهم‌ریختگی انواع معیارهای Precision ، Recall ، f1-score و Accuracy را محاسبه نمود



شکل ۳: ماتریس درهم‌ریختگی برای پیش‌بینی حاملگی مثبت و منفی

در ادامه برای نمونه‌های تست که در گام دوم دریافت شده را روی مدل آموزش دیده اجرا نموده‌ایم.

۵-نتیجه‌گیری

با توجه به نتایج حاصل از این مقاله، می توان نتیجه گیری کرد که استفاده از تکنیک های هوش مصنوعی برای پیش بینی آسیب DNA و کیفیت جنین در لقاح مصنوعی، می‌تواند بهبود قابل توجهی در عملکرد و دقت پیش‌بینی داشته باشد. این روش می تواند به عنوان یک ابزار موثر برای کمک به پزشکان و متخصصان در انجام لقاح مصنوعی و پیش بینی موفقیت آن، مورد استفاده قرار گیرد. این مقاله نشان دهنده اهمیت و کارایی تکنیک های هوش مصنوعی در پیش بینی و تخمین داده های پزشکی است و می تواند به توسعه روش های بهبود یافته در این زمینه کمک کند.

منابع

[1] Armina, Roslan, et al. "A review on missing value estimation using imputation algorithm." Journal of Physics: Conference Series. Vol. 892. No. 1. IOP Publishing, 2017..

[2] Troyanskaya, Olga, et al. "Missing value estimation methods for DNA microarrays." Bioinformatics 17.6 (2001): 520-525.

[3] . Choong, Miew Keen, Maurice Charbit, and Hong Yan. "Autoregressive-model-based missing value estimation for DNA microarray time series data." IEEE Transactions on information technology in biomedicine 13.1 (2009): 131-137.

[4] Keerin, Phimmarin, and Tossapon Boongoen. "Improved knn imputation for missing values in gene expression data." Computers, Materials and Continua 70.2 (2021): 4009-4025.

[5] Abnane, Ibtissam, et al. "Analogy software effort estimation using ensemble KNN imputation." 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, 2019.

[6] Boquet, Guillem, et al. "Missing data in traffic estimation: A variational autoencoder imputation method." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019

[7] . Li, Yifeng, and Alioune Ngom. "The non-negative matrix factorization toolbox for biological data mining." Source code for biology and medicine 8.1 (2013): 1-15.

[8] He, Xiangnan, et al. "Fast matrix factorization with nonuniform weights on missing data." IEEE transactions on neural networks and learning systems 31.8 (2019): 2791-2804.

[9] Alonso, Sergio, et al. "A learning procedure to estimate missing values in fuzzy preference relations based on additive consistency." Modeling Decisions for Artificial Intelligence: First International Conference, MDAI 2004, Barcelona, Spain, August 2-4, 2004. Proceedings 1. Springer Berlin Heidelberg, 2004.

[10] Leke, Collins, and Tshilidzi Marwala. "Missing data estimation in high-dimensional datasets: A swarm intelligence-deep neural network approach." Advances in Swarm Intelligence: 7th International Conference, ICSI 2016, Bali, Indonesia, June 25-30, 2016, Proceedings, Part I 7. Springer International Publishing, 2016.

[11] Kim, Hyunsoo, Gene H. Golub, and Haesun Park. "Missing value estimation for DNA microarray gene expression data: local least squares imputation." Bioinformatics 21.2 (2005): 187-198.

[12] Bloch, Isabelle. "Information combination operators for data fusion: A comparative review with classification." IEEE Transactions on systems, man, and cybernetics-Part A: systems and humans 26.1 (1996): 52-67.

[13] Polikar, Robi. "Ensemble based systems in decision making." IEEE Circuits and systems magazine 6.3 (2006): 21-45.

[14] Zhang, Cha, and Yunqian Ma, eds. Ensemble machine learning: methods and applications. Springer Science & Business Media, 2012.

[15] Freund, Yoav, and Robert E. Schapire. "A desicion-theoretic generalization of on-line learning and an application to boosting." European conference on computational learning theory. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.

[16] Nakashima, Hayato, Ismail Arai, and Kazutoshi Fujikawa. "Passenger counter based on random forest regressor using drive recorder and sensors in buses." 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2019.

[17] You, Dongfang, et al. "Identification of genetic features associated with fine particulate matter (PM2. 5) modulated DNA damage using improved random forest analysis." Gene 740 (2020): 144570.

[18] Graw, J. H., W. T. Wood, and B. J. Phrampus. "Predicting global marine sediment density using the random forest regressor machine learning algorithm." Journal of Geophysical Research: Solid Earth 126.1 (2021): e2020JB020135.

[19] El Mrabet, Zakaria, et al. "Random Forest regressor-based approach for detecting fault location and duration in power systems." Sensors 22.2 (2022): 458.

[20] Armi, Laleh, Elham Abbasi, and Jamal Zarepour-Ahmadabadi. "Mixture of ELM based experts with trainable gating network." arXiv preprint arXiv:2105.11706 (2021).

[21] Armi, Laleh, Elham Abbasi, and Jamal Zarepour-Ahmadabadi. "Texture images classification using improved local quinary pattern and mixture of ELM-based experts." Neural Computing and Applications (2021): 1-24.

[22] Armi, Laleh, and Hossein Ebrahimpour-Komleh. "Classification of Skin Cancer With Using Color-ILQP and MEETG." 2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA). IEEE, 2023.

[23] Armi, Laleh, and Elham Abbasi. "Tree Bark Classification using Color-improved Local Quinary Pattern and Stacked MEETG." Journal of AI and Data Mining 11.3 (2023): 391-405.

.