



تشخیص میزان بدخیمی غدد لنفاوی با استفاده از روشهای مختلف یادگیری ماشین

بهار جهدوری^{۱*}، سیده مبینا حسینی راد^۲، سامان راجبی^۳، نادر وحدانی مناف^۴

۱- دانشجوی کارشناسی مهندسی پزشکی، موسسه آموزش عالی سراج

۲- دانشجوی کارشناسی مهندسی پزشکی، موسسه آموزش عالی سراج

۳- هیات علمی گروه مهندسی پزشکی، موسسه آموزش عالی سراج

۴- هیات علمی گروه مهندسی پزشکی، موسسه آموزش عالی سراج

*Baharjahdevari@gmail.com

ارسال: اردیبهشت ماه ۱۴۰۳ پذیرش: مرداد ماه ۱۴۰۳

چکیده

در سالیان اخیر تکنیک های یادگیری ماشین به طور فزایندهای در حیطه مراقبت های بهداشتی، به دلیل توانایی آنها در تجزیه و تحلیل حجم عظیمی از داده و کمک به تشخیص های دقیق، استفاده می شوند. در زمینه لنفوگرافی، طبقه بندی گره های لنفی نقش حیاتی در تشخیص بیماری و برنامه ریزی درمان دارد بنابراین خطاها در تفسیر تصاویر لنفوگرافی می تواند منجر به اشتباهات تشخیصی و نتایج نامطلوب برای روند درمان بیمار شود. مجموعه داده مورد بررسی در این مطالعه شامل ۱۸ ویژگی بارز موجود در تصاویر لنفوگرافی است که در این مطالعه به بررسی کاربرد تکنیک های یادگیری ماشین، به خصوص برای طبقه بندی این تصاویر لنفوگرافی پرداخته شده است. جهت طبقه بندی دقیق انواع مختلف گره های لنف بر اساس ویژگی های آن ها، الگوریتم های کلاسبندی KNN، SVM و MLP مورد بررسی قرار گرفته شد که کلاسبند SVM با نرخ صحیح کلاسبندی ۸۷/۱٪ بهترین نتیجه گزارش شد. همچنین الگوریتم بهینه سازی PSO جهت انتخاب بهترین ویژگی ها مورد استفاده قرار گرفته و با استفاده از ۵ ویژگی برتر انتخاب شده کلاسبند SVM در کوتاه ترین زمان با دقت ۹۶/۶٪ به عنوان بهترین الگوریتم معرفی شد.

واژگان کلیدی: انتخاب ویژگی، بهینه سازی ازدحام ذرات، شبکه عصبی، لنفوگرافی، یادگیری ماشین.

۱- مقدمه

سیستم عروق لنفاوی یک مجرای یک طرفه است که مایع شریانی بینا بینی فیلتر شده و متابولیت های بافتی را به گردش خون بر می گرداند. گره لنفاوی عضوی بیضی شکل از این سیستم است که به طور گسترده در سرتاسر بدن توزیع شده و با افزایش اندازه به دلایل مختلف تمایل دارد که نشان دهنده عوارض سلامتی باشد. این عوارض می توانند بیماری های بی اهمیت و بیماری هایی که سلامت انسان را به طور جدی تهدید می کنند: از جمله ورم لنفی یا لنف ادم، متاستاز یا فراگستری سرطان و اختلالات التهابی مختلف، باشند [۱-۲].

در سال ۲۰۱۸، شیوع تخمینی متاستاز سرطان - برای ۶ نوع سرطان رایج - ۶۲۳۴۰۵ بود. در این پیش بینی احتمال افزایش شیوع متاستاز سرطان برای این ۶ نوع در سال ۲۰۲۵ به ۶۹۳۴۵۲، مشاهده می شود. همینطور از هر ۵ زن که از سرطان سینه جان سالم به در

می برد، یک نفر دچار لنف‌ادم می شود. در سرطان سر و گردن، عوارض لنفاوی و بافت نرم می تواند در طول ۱۸ ماه اول پس از درمان ایجاد شود و بیش از ۹۰ درصد بیماران نوعی لنف‌ادم داخلی، خارجی یا ترکیبی را تجربه می کنند. در یک مطالعه اخیر، ۳۷ درصد از زنان تحت درمان سرطان زنان، شواهد قابل اندازه گیری لنف‌ادم را طی ۱۲ ماه پس از درمان داشتند. تشخیص درست و به موقع این اختلالات آسان نیست، برای مثال پاتولوژی لنف‌ادم می تواند با سایر تظاهرات بالینی (به عنوان مثال، ادم محیطی ناشی از نارسایی وریدی) اشتباه گرفته شود [۷-۳].

از آنجایی که سرطان عامل اصلی مرگ و میر در سراسر جهان است، لنفوگرافی^۱ از پیشرفت های قابل توجه اخیر دانش پزشکی محسوب می شود. لنفوگرافی یک تکنیک تصویربرداری پزشکی است که در آن رادیو ماده حاجب تزریق می شود و سپس یک عکس اشعه ایکس برای تجسم ساختارهای سیستم لنفاوی گرفته می شود و علاوه بر این از وضعیت موجود غدد لنفاوی نیز می توانند احتمال بروز سرطان را نشان دهند. این داده ها برای تصمیم گیری در مورد اینکه آیا جزئیات بالینی به دست آمده از لنفوگرافی مربوط به یک یافته طبیعی یا غیر طبیعی می باشند بسیار ضروری است. از این رو کلاس بندی مناسب و تعیین ویژگی های معتبر می تواند فرآیند پیش بینی بیماری را ساده کرده و اقدامات بازدارنده را برانگیزد. هوش مصنوعی این کار پیچیده را با آموزش دادن ماشین آلات هوشمند به اندازه انسان ممکن کرده تا الگوهایی را در شرایط مختلف محیطی بشناسند و آنها را در دسته های دلخواه به روشی ساده و قابل اعتماد طبقه بندی کنند. چنین شاخه ای از هوش مصنوعی به عنوان تشخیص الگو شناخته می شود. تشخیص الگو راه حل بسیاری از مشکلات را که در مبحث طبقه بندی قرار دارند، فراهم می کند [۲ و ۸].

۱-۱- کارهای پیشین

علی البطینه^۲ و همکارانش در سال ۲۰۲۲ جهت پیشنهاد روشی برای شناسایی بیماری های قلبی با استفاده از مجموعه داده بیماری قلبی کیولند^۳ مطالعه ای انجام داده اند که این داده ها شامل ۱۳ ویژگی از جمله سن، جنسیت، نوع درد در ناحیه سینه، فشار دیاستولی و غیره از ۳۰۳ نمونه می شوند. هدف از این مطالعه بررسی اثربخشی طبقه بندی کننده های MLP^۴ آموزش دیده با PSO^۵ برای توسعه طبقه بندی باینری مدل های که می توانند به عنوان تشخیص بیماری قلبی استفاده شوند بود. در نتیجه روش پیشنهادی MLP+PSO از الگوریتم های دیگر بهتر عمل کرد و با دقت ۸۴/۶٪ از مابقی الگوریتم ها پیشی گرفت. همینطور در سال ۲۰۰۶ کمال پلات^۶ و همکارش سیستمی به نام Fuzzy-AIRS^۷ به عنوان طبقه بندی کننده در تشخیص بیماری های لنفاوی مورد بررسی قرار دادند. این طبقه بندی با داده های بیماری های لنفاوی گرفته شده از دانشگاه کالیفرنیا در ایروین^۸ با استفاده از روش اعتبار سنجی متقابل ۱۰ برابر انجام شد. مجموعه داده این بررسی نیز از مرکز پزشکی دانشگاه و موسسه انکولوژی لیوبلیانا در یوگسلاوی در مجموع ۱۴۸ نمونه با ۱۸ ویژگی جمع آوری شده است و در نتایج این مطالعه ذکر شده است که سیستم طبقه بندی AIRS به دقت ۸۳٪/۱ رسیده در حالی که مجموعه داده های بیماری های لنفاوی طبقه بندی شده با Fuzzy-AIRS با دقت ۹۰٪ گزارش شدند [۹-۱۰].

هدف از انجام تحقیقات در این مقاله ارائه کلاس بند کارآمد برای تشخیص اتوماتیک وضعیت یک غده بر اساس ویژگی های دیگر با استفاده از هوش مصنوعی خواهد بود. بخش روش کار، روش های طبقه بندی ماشین بردار پشتیبان^۹ SVM، همسایگی نزدیک ترین همسایه ها^{۱۰} KNN و شبکه عصبی چندلایه پرسپترون MLP با شرح دقیق اهداف هر الگوریتم در ایجاد مرزهای تصمیم، طبقه

¹ Lymphography

² Ali Al Bataineh

³ Cleveland

⁴ Multi-layer perceptron

⁵ Particle Swarm Optimization

⁶ Kemal Polat

⁷ AIRS: Artificial immune recognition system

⁸ University of California at Irvine (UCI) Machine Learning Repository

⁹ Support Vector Machine

¹⁰ K-Nearest Neighbors

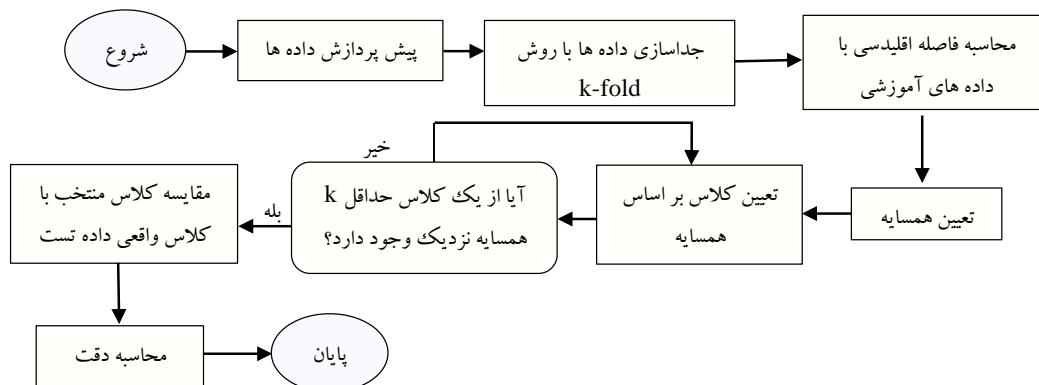
بندی همسایگان نزدیک و یادگیری شبکه عصبی توضیح داده می شود. بخش نتایج شبیه سازی به ارزیابی و تحلیل مقایسه ای عملکرد هر الگوریتم با استفاده از معیار های نرخ درست کلاس بندی و زمان اجرای هر کلاس بند ها در اثر تغییر یک سری ویژگی ها می پردازد و سپس با استفاده از روش بهینه سازی PSO برای بهبود فرآیند طبقه بندی تحت انتخاب ویژگی قرار می گیرند و بهترین نوع روش طبقه بندی برای بهبود برنامه های یادگیری ماشین در بخش بحث ارائه داده خواهد شد.

۲- تئوری روشهای کلاسبندی

در طی سالیان متمادی، مجموعه ای از روش های کلاسبندی جهت شناسایی الگوهای موجود در داده ها و طبقه بندی آنها بر اساس آن الگوهای شناسایی شده، کشف شده است. تصمیم گیری در مورد اینکه کدام روش برای طبقه بندی داده ها با کمترین زمان و بهترین نتایج، مناسب تر می باشد، همیشه دشوار بوده است. از این رو داشتن شناخت کلی از برخی روش ها می تواند به مقایسه آن ها و تصمیم گیری بسیار کمک کند.

۲-۱- کلاس بندی نزدیک ترین همسایه (KNN)

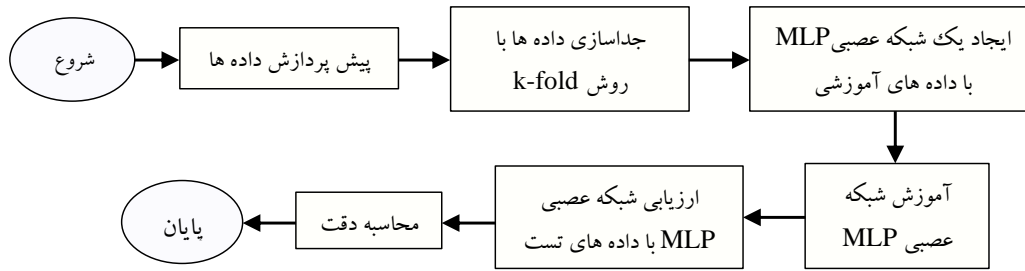
این کلاس بند یکی از ساده ترین کلاس بند های یادگیری ماشینی است که در آن یک نمونه ذره با استفاده از رای اکثریت همسایه هایش طبقه بندی می شود و سپس ذره به کلاسی که در بین K نزدیک ترین همسایه هایش رایج ترین است، اختصاص می یابد. در اینجا همسایگان از مجموعه ای از اشیاء گرفته شده اند که طبقه بندی صحیح آنها مشخص است. این همسایگان را می توان به عنوان یک مجموعه آموزشی برای این الگوریتم فرض کرد، اگرچه هیچ مرحله آموزشی صریحی لازم نیست. یادگیری در این مدل مبتنی بر ذخیره تمام نمونه های آموزشی است که مربوط به نقاطی در فضای اقلیدسی Π بعدی به همراه برجسب های کلاس آنها است و طبقه بندی تا رسیدن یک نمونه جدید به تاخیر می افتد. در شکل ۱، فلوچارت روش طبقه بندی این کلاسبند آمده است [۸، ۱۱-۱۲].



شکل ۱- فلو چارت روش طبقه بندی کننده KNN

۲-۲- کلاس بندی پرسپترون چند لایه (MLP)

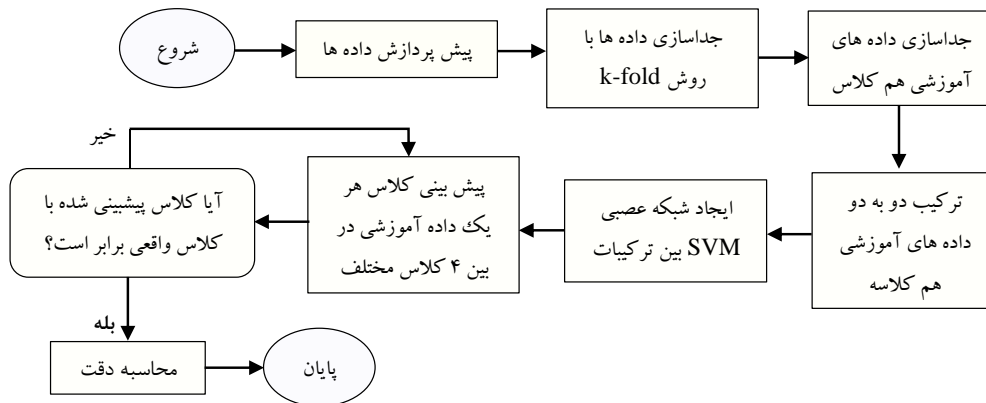
پرسپترون چند لایه شاید محبوب ترین معماری شبکه ای باشد که امروزه استفاده می شود و به طور گسترده برای طبقه بندی الگوها استفاده می شود. MLP از تعداد زیادی نورون به هم پیوسته تشکیل شده است که به طور موازی برای حل یک مشکل خاص کار می کنند. معماری اصلی یک شبکه MLP شامل سیگنال هایی است که به صورت متوالی در لایه های مختلف از لایه ورودی به لایه خروجی جریان می یابد. بین لایه ورودی و لایه خروجی، لایه های پنهان قرار دارند. سیگنال ها به گره های لایه ورودی منتقل می شوند، از طریق لایه های پنهان به جلو حرکت می کنند و در نهایت از لایه خروجی خارج می شوند. هر گره، در لایه پنهان، خروجی گره قبل را از لایه ورودی از طریق اتصال وزن دریافت می کند و سپس پاسخ مربوطه را تولید می کند که به لایه خروجی ارسال می شود. در واقع، هر گره یک مجموع وزنی را انجام می دهد که توسط یک تابع غیر خطی منتقل می شود. در شکل ۲، فلوچارت روش طبقه بندی این کلاسبند آمده است اکثر برنامه های کاربردی شبکه عصبی شامل MLP ها هستند [۱۳-۱۵].



شکل ۲- فلو چارت روش طبقه بندی کننده MLP

۳-۲- کلاس بند ماشین بردار پشتیبانی (SVM)

SVM طبقه بندی را با ساخت یک ابر صفحه N بعدی انجام می دهد که داده ها را به طور بهینه به دو دسته جدا می کند. همانطور که در شکل ۳ مشاهده می شود، این کلاس بند مجموعه ای از داده های ورودی را می گیرد و کلاس ممکن را پیش بینی می کند. برای هر ورودی از بین دو کلاس ممکن، ورودی عضوی از آن است که آن را به یک طبقه بندی خطی باینری غیر احتمالی تبدیل می کند و در واقع الگوریتم آموزشی SVM مدلی را ایجاد می کند که نمونه های جدیدی را به یک دسته یا دسته دیگر اختصاص می دهد. از آن جایی که مجموعه داده لئوگرافی متشکل از ۴ کلاس متمایز است، از SVM دو کلاسه سنتی در طبقه بندی چند کلاسه استفاده می شود که در آن داده هایی که باید طبقه بندی شوند ممکن است به هر یک از کلاس ها در میان تعدادی از کلاس ها تعلق داشته باشند [۸].



شکل ۳- فلو چارت روش طبقه بندی کننده SVM چهار کلاسه

۳- نتایج شبیه سازی

در این بخش، عملکرد طبقه بندی کننده های انتخاب شده روی مجموعه داده لئوگرافی ارائه می شود. علاوه بر این، با استفاده از تجزیه و تحلیل آماری شکل های مختلف طبقه بندی کننده ها، بهترین روش و نتایج آن گزارش خواهد شد. تمامی نتایج گزارش شده تحت رم ۸ گیگابایتی و پردازش گر ۲/۳۰ GHz انجام گرفته شده است.

۳-۱- مجموعه داده لئوگرافی

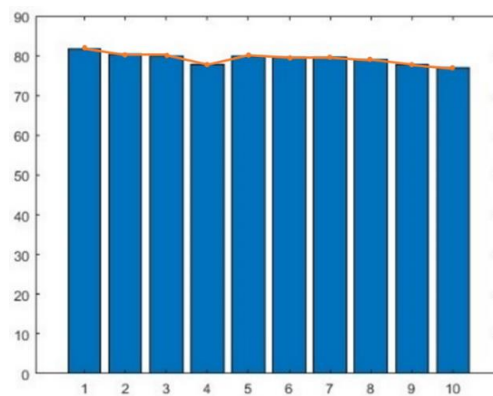
مجموعه داده لئوگرافی مورد استفاده در این مقاله از مرکز پزشکی دانشگاه، موسسه انکولوژی، لیوبلیانا، یوگسلاوی به دست آمده است. این مجموعه ۱۴۸ مورد را برای آموزش طبقه بندی کننده فراهم می کند. جزئیات ویژگی ها، مقادیر احتمالی آنها و شناسه ویژگی مرتبط به وضوح در جدول ۱ فهرست شده است. در این مجموعه داده ستون اول از یک کلاس هدف با ۴ مقدار متمایز تشکیل شده است که می تواند چهار مقدار مجزا داشته باشد و تعداد ویژگی های پیش بینی کننده تا هجده مجموع می باشد برای آموزش طبقه بندی کننده ها استفاده می شوند.

جدول ۱- مشخصات مجموعه داده لنفوجرافی [۱۶]

شماره ویژگی	ویژگی	حالات ممکن	محدوده مقادیر
۱	کلاس	خوش خیم، سرطان بدخیم، لنف بدخیم، فیروز	۱-۴
۲	غدد لنفاوی	نرمال، قوس دار، تغییر شکل یافته، جابجا شده	۱-۴
۳	بلوک آوران	بله، خیر	۱،۲
۴	بلوک لنفاوی C	بله، خیر	۱،۲
۵	بلوک لنفاوی S	بله، خیر	۱،۲
۶	بای پس شدن	بله، خیر	۱،۲
۷	تراوش	بله، خیر	۱،۲
۸	بازسازی شدن	بله، خیر	۱،۲
۹	جذب اولیه	بله، خیر	۱،۲
۱۰	غدد لنفاوی کاهش یافته	۰-۳	۰-۳
۱۱	غدد لنفاوی بزرگ شده	۱-۴	۱-۴
۱۲	تغییرات در لنف	لوبیا شکل، بیضی، گرد	۱-۳
۱۳	نقص در غدد	خیر، حفره، حفره مرزی، حفره مرکزی	۱-۴
۱۴	تغییرات در غدد	خیر، حفره، حفره مرزی، حفره مرکزی	۱-۴
۱۵	تغییرات در ساختار	خیر، دانه دار، قطره مانند، درشت، رقیق شده، مشبک، کنده شده، کمرنگ	۱-۸
۱۶	صورت های خاص	خیر، جام، وزیکول	۱-۳
۱۷	جا به جایی	بله، خیر	۱،۲
۱۸	حالت استثنا	بله، خیر	۱،۲
۱۹	تعداد غدد لنفاوی	۰-۹،۱۰-۱۹،۲۰-۲۹،۳۰-۳۹،۴۰-۴۹،۵۰-۵۹،۶۰-۶۹،۷۰=<	۱-۸

۳-۲- نتایج کلاسبندی با استفاده از الگوریتم نزدیکترین همسایه

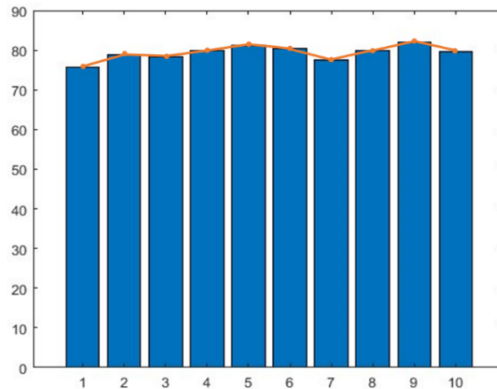
مطابق با توضیحات ارائه شده در بخش ۲-۱، یک کلاسبند KNN با مدل ارزیابی K-fold به ازای ۱۰ بسته و شرط تعداد همسایه به ازای ۲ تا ۲۰ عدد زوج تربیتی طراحی شده است و اثرات تغییر تعداد همسایه ها بر عملکرد تشخیص آن با ده بار اجرای کلاسبند و میانگین گیری از ۱۰ مقدار CCR^۱ برای هر مقدار K، طبق شکل ۴ ارزیابی گردید. در نمودار شکل ۴ روند تغییرات CCR، نزولی مشاهده می شود. به طوری که در مدت زمان ۱۰/۸ ثانیه، با شرط حداقل دو همسایه کلاسبند KNN به بیشترین دقت یعنی ۸۱/۸٪ رسیده و کمترین دقت برای ستون دهم یعنی شرط بیست همسایه برای تعیین کلاس مشاهده می شود.



شکل ۴- نمودار تغییرات CCR بر حسب تعداد زوج همسایگی

^۱ نرخ صحیح کلاسبندی: Correct Classification Rate

با توجه به شکل ۴ به دلیل اینکه نرخ صحیح کلاسبندی در دفعاتی که شرط تعداد همسایه ها کمتر از ۱۰، بالاتر بود؛ در ادامه کلاسبندی KNN به ازای تعداد ۱ تا ۱۰ عدد همسایه در شکل ۵ نشان داده شده است.

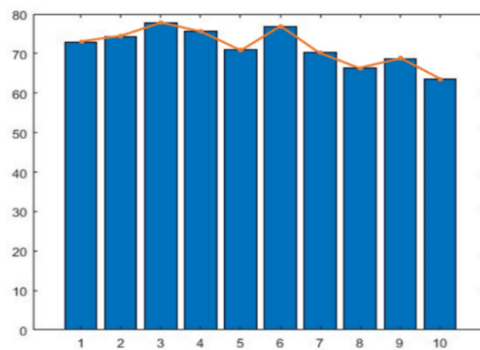


شکل ۵- نمودار تغییرات CCR بر حسب تعداد همسایگی

کلاسبندی KNN در طبقه بندی مجموعه داده لئوگرافی بهترین نرخ صحیح کلاس بندی را به ازای ۹ همسایه در ۱۰/۶ ثانیه گزارش می کند که به دقت ۸۱/۹٪ رسیده است. اگرچه الگوریتم KNN یک روش ساده، تحت نظارت و پرکار برد است اما نمی تواند نتیجه رضایت بخشی از عملکردش در بعضی از برنامه ها به عمل بیاورد چرا که به عنوان الگوریتم، پیچیدگی محاسباتی بالا و سرعت به نسبت پایینی دارد.

۳-۳- نتایج شبیه سازی روش MLP

طبق توضیحات ارائه شده در بخش های قبل، با استفاده از متد ارزیابی K-fold، داده های آموزشی و تست به ازای ۱۰ بسته مشخص شده و با داده های آموزشی شبکه های عصبی MLP یک لایه و دو لایه طراحی شده است. در این شبکه های عصبی ۸۰ درصد داده ها برای آموزش شبکه، ۱۵ درصد داده ها برای اعتبار سنجی و سپس ۵ درصد داده ها برای ارزیابی شبکه انتخاب شدند. از ویژگی های روش MLP می توان به نوسانات زیاد بین نتایج کلاس بندی اشاره کرد؛ به همین دلیل نتیجه هر بخش به تنهایی گزارش نشده و همه نرخ های صحیح کلاس بندی با میانگین گیری از ۳۰ بار اجرا مورد بررسی قرار گرفته اند و نتایج و نرخ صحیح کلاسبندی آن به ازای نوروں های متفاوت در لایه پنهان مطابق شکل ۶ گزارش شده است.



شکل ۶- نمودار هیستوگرام تغییرات CCR در یک لایه پنهان

در شکل ۶، شبکه عصبی MLP با یک لایه پنهان به ازای تعداد نوروں های ۱ تا ۱۰ در مدت زمان ۲۵۴/۷ ثانیه شبیه سازی شده و با توجه به نتایج بدست آمده بهترین نتیجه با ۷۵/۵٪ نرخ صحیح کلاسبندی با ۹ عدد نوروں به دست آمده است.

۳-۴- نتایج کلاسبندی با الگوریتم ماشین بردار پشتیبان

مطابق با توضیحات ارائه شده در بخش ۲-۳ و شکل ۳، یک کلاسبندی SVM با مدل ارزیابی K-fold طراحی شده است. در این بخش اثرات تغییر تعداد داده های آموزشی و تست با تغییر تعداد بسته های روش K-fold بر نرخ صحیح کلاسبندی و مدت زمان اجرا ارزیابی گردید و نتایج حاصل از این تغییرات با میانگین گیری از کلیه حالات در جدول ۲ آمده است.

جدول ۲- CCR و زمان اجرا به ازای تغییر تعداد بسته های K-fold

تعداد بسته	۱۰	۹	۸	۷	۶
CCR(%)	۸۷/۱	۸۵/۱	۸۴/۳	۸۳/۱	۷۹
زمان اجرا (s)	۱/۵	۱/۴	۱/۲	۱/۲	۱/۵

با توجه به جدول ۲، می توان نتیجه گرفت که برای کلاسنبدی مجموعه لنفوگرافی با روش SVM، کمتر کردن تعداد بسته ها در مدل ارزیابی K-fold جهت جدا کردن داده های تست و آموزشی باعث افت نرخ صحیح کلاسنبدی می شود و با اینکه SVM روش کلاسنبدی نسبتاً سریعی است، این تغییر پیشرفت چشمگیری در زمان اجرا ندارد. در این صورت بهترین عملکرد هنگامی صورت می گیرد که با استفاده از متد K-fold داده ها با ۱۰ بسته به بخش های آموزشی و تست جدا شوند.

۴- انتخاب ویژگی

در بسیاری از حالت های طبقه بندی الگو، تعداد بیشتر ویژگی های مورد استفاده لزوماً به دقت طبقه بندی بالاتر موجب نمی شود. در برخی مواردی که عملکرد الگوریتم هایی مختص به سرعت و دقت پیش بینی توصیف داده ها هستند؛ می تواند کاهش یابد. بنابراین هدف از انتخاب ویژگی کاهش حداکثر تعداد ویژگی های نامرتب با حفظ دقت طبقه بندی قابل قبول است. انتخاب ویژگی در طبقه بندی الگو، تجزیه و تحلیل داده ها، پردازش داده های پزشکی و غیره اهمیت قابل توجهی دارد. چندین روش برای انجام انتخاب ویژگی در آموزش و آزمایش داده ها استفاده می شود در این مقاله به منظور بهینه سازی ویژگی ها از الگوریتم بهینه سازی ازدحام ذرات (PSO) استفاده می شود [۱۷].

۴-۱- الگوریتم بهینه سازی PSO باینری

الگوریتم بهینه سازی ازدحام ذرات (PSO) یک الگوریتم جستجوی مبتنی بر جمعیت است که بر اساس شبیه سازی تجمع پرندگان و توسط رینولدز^۱ توسعه یافته است. مشابه سایر الگوریتم های محاسباتی تکاملی، PSO یک الگوریتم جستجو مبتنی بر جمعیت است که از مدل اشتراک گذاری اجتماعی اطلاعات استفاده می کند. در PSO، افراد (به نام ذرات^۲) در فضای جستجوی فوق بعدی پرواز می کنند. تغییرات در موقعیت ذرات در فضای جستجو بر اساس تمایل افراد به تقلید از موفقیت دیگران در جمعیت^۳ است. بنابراین رفتار جستجوی یک ذره بر این اساس تحت تأثیر رفتار سایر ذرات درون ازدحام قرار می گیرد. به عبارت دیگر، ذرات در یک فضای جستجوی چند بعدی به پرواز در می آیند که در آن موقعیت هر ذره مطابق با تجربه خود و همسایگانش تنظیم می شود [۱۸]. در بکارگیری الگوریتم بهینه سازی PSO جهت انتخاب ویژگی، متغیرهای باینری در نظر گرفته می شوند. بطوریکه بیت صفر به معنی حذف ویژگی متناظر و بیت یک به معنی در نظر گرفتن ویژگی متناظر آن خواهد بود. در چنین حالتی نرخ صحیح کلاسنبدی به ازای ویژگی های در نظر گرفته شده به عنوان تابع ارزیابی PSO در نظر گرفته خواهد شد.

۴-۱-۱- نتایج انتخاب ویژگی با استفاده از الگوریتم PSO

الگوریتم بهینه سازی ازدحام ذرات (PSO) برای انتخاب ویژگی های متناسب دارای ۳ متغیر می باشد که شامل تعداد ویژگی های مجموعه داده مورد استفاده، تعداد ذرات (جمعیت) و تعداد دفعات اجرا می شوند. همچنین نسبت ضرایب فاصله یعنی $C1, C2, W$ و Wref نیز در نتیجه نهایی نقش بزرگی را ایفا می کنند. تحلیل و آزمایشات متعدد با پارامترهای مختلف می تواند به بهبود الگوریتم برای داده ها و اهداف خاص کمک کند. بنابراین ابتدا برای تعیین پیکر بندی بهینه به ازای ۱۸ ویژگی مجموعه داده لنفوگرافی، برنامه به ازای تغییرات دو متغیر تعداد دفعات تکرار و تعداد ذره به ترتیب در جدول ۳ و جدول ۴، اجرا شده و نتایج ثبت شده اند.

¹ Reynolds

² Particles

³ swarm

جدول ۲- الگوریتم بهینه سازی PSO به ازای تغییر دفعات تکرار

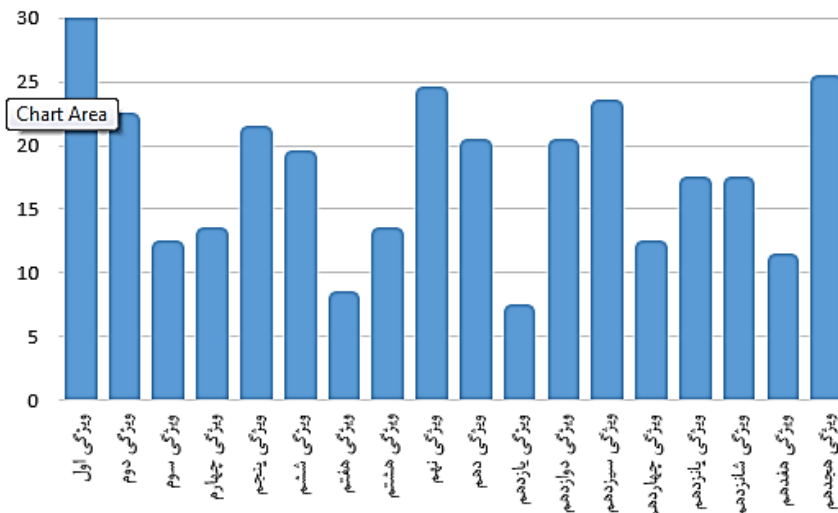
تعداد ذرات	۴۰	۴۰	۴۰
تعداد دفعات اجرا	۴۰۰	۲۰۰	۱۰۰
بهترین ذره	۳,۳۳۳	۳,۳۳۳	۳,۳۳۳
زمان اجرا (s)	۴۹۸۹/۸	۲۸۴۰/۲	۱۳۷۴/۱

در جدول ۳، نتایج الگوریتم انتخاب ویژگی PSO، بهترین نتیجه با مقدار حداکثر ۴۰ ذره، با کمترین زمان اجرا به ازای ۱۰۰ بار تکرار برنامه به دست آمده است. به نظر می رسد که تعداد دفعات تکرار تأثیری در بهترین نتیجه نداشته و فقط باعث تغییر در زمان عملکرد شده است و این نتیجه نشان دهنده این است که افزایش تعداد تکرارها در الگوریتم بهینه سازی PSO برای مجموعه داده لئوگرافی تغییری در خروجی بهترین ذره ایجاد نمی کند و منجر به هم گرایی بهتر و بهبود نتایج نمی شود. علاوه بر این تأثیر تعداد ذرات بر عملکرد الگوریتم به ازای حداکثر ۲۰۰ تکرار را می توان با مقایسه نتایج با مقایسه نتایج جدول ۴ مشاهده کرد.

جدول ۳- الگوریتم بهینه سازی PSO به ازای تغییر تعداد ذرات

تعداد ذرات	۲۰	۴۰	۶۰
تعداد دفعات اجرا	۲۰۰	۲۰۰	۲۰۰
بهترین ذره	۳/۳۳	۳/۳۳	۳/۳۳
زمان اجرا (s)	۱۳۷۴/۱	۲۸۴۰/۲	۴۴۱۹/۲

بر اساس جدول ۴ مشاهده می شود که با وجود زمان های عملکرد متفاوت برای تعداد مختلف ذرات و ۲۰۰ بار تکرار برنامه، الگوریتم بهینه سازی بطور پایدار بهترین نتیجه را ارائه می دهد. نتایج هر دو جدول ۳ و ۴ نشان می دهد که الگوریتم موفق به شناسایی مجموعه بهینه ویژگی ها برای مجموعه داده لئوگرافی شده است. با بررسی نتایج الگوریتم PSO در جدول ۳ و ۴، مشخص شد که برخی از ویژگی ها به طور مداوم در چندین اجرا انتخاب شده اند. این ویژگی ها می توانند به عنوان بسیار مهم در تمایز متغیر هدف در نظر گرفت شوند، زیرا انتخاب مکرر آنها نشان دهنده اهمیت آنها در مدل پیش بینی است. در شکل ۷، نمودار تعداد انتخاب شدن ویژگی ها آمده است که در انتخاب مهم ترین ویژگی ها کمک می کند زیرا انتخاب مکرر آنها نشان دهنده اهمیت آنها در مدل پیش بینی است. نتایج انتخاب ویژگی توسط الگوریتم PSO نشان می دهد که یک ترجیح ثابت برای اولین ویژگی وجود دارد که در همه ۳۰ اجرای الگوریتم انتخاب شده است. از سوی دیگر، ویژگی یازدهم به عنوان کمترین انتخاب شده، تنها در ۷ بار از مجموع ۳۰ اجرا انتخاب شده است.



شکل ۷- نمودار تراکم انتخاب هر ویژگی در ۳۰ حالت مختلف

انتخاب مکرر ویژگی اول نقش مهم آن را در تمایز متغیر هدف تأکید می کند، در حالی که انتخاب کم فراوان ویژگی یازدهم نشان دهنده تأثیر کمتر آن بر توان پیش بینی مدل است. علاوه بر این، تغییرپذیری در فراوانی انتخاب شدن ویژگی های مختلف، پایداری

الگوریتم و تطبیق آن با تنظیمات مختلف را نشان می دهد. ویژگی هایی با فراوانی انتخاب ممکن است حساسیت کم آن ها نسبت به تغییرات در پارامترهای الگوریتم را نشان دهند، که به معنای وجود یک تعادل بین اهمیت آنها و پایداری در مدل است.

۵- نتیجه گیری

در این مقاله، روش های SVM، KNN و MLP جهت طبقه بندی مجموعه داده لنفوگرافی ارزیابی شده است. بهترین نتیجه به دست آمده از روش MLP با یک لایه پنهان ۹ نورونی به دست آمد ولی این روش برای طبقه بندی مناسب نبوده زیرا به کمترین دقت یعنی ۷۵/۵٪ رسیده و به دلیل ناپایداری شبکه عصبی به منظور گزارش نتیجه مطلوب کلاسیک باید ۳۰ بار اجرا شده و نتایج به صورت میانگین گزارش می شدند که باعث بالا رفتن حجم محاسبات و زمان اجرا شد. روش KNN در ۱۰/۶ ثانیه با شرط ۹ همسایه به دقت ۸۱/۹٪ رسیده و از روش MLP سریع تر بوده و به دقت بالاتری رسیده است. در نهایت روش SVM در مدت زمان ۱/۵ ثانیه دقت ۸۷/۱٪ گزارش کرده و نسبت به دو روش قبل در مدت زمان کمتر به دقت بالایی رسیده است. همچنین در ادامه با استفاده از الگوریتم PSO ویژگی هایی که موجب بهبود بخشی به عملکرد کلاسیک و همچنین ویژگی هایی که مناسب کلاسیک نبوده اند، مشخص شده اند. بنابراین در جدول ۵ بهترین حالت هر روش به ازای ویژگی های منتخب گزارش شده است.

جدول ۴- روش های کلاسیک به ازای همه ویژگی و ۵ ویژگی منتخب

روش کلاسیک	SVM	KNN	MLP
CCR با ۱۸ ویژگی (%)	۸۷/۱	۸۱/۹	۷۵/۵
زمان اجرا (s)	۱/۵	۱۰/۶	۲۵۴/۷
CCR با ۵ ویژگی برتر (%)	۹۶/۶	۸۰	۹۵/۷
زمان اجرا (s)	۱/۵	۴/۴	۱۰۱/۳

با ۵ ویژگی شناسایی شده توسط الگوریتم PSO هر ۳ روش بار دیگر ارزیابی شده است. بطوری که در روش SVM اجرای کلاسیک با ۵ ویژگی منتخب موجب ۹٪ افزایش دقت گردید. در روش MLP این تغییر موجب ۱۸/۸٪ افزایش دقت و کاهش زمان اجرا و در روش KNN موجب کاهش ۱/۹٪ دقت ولی کاهش زمان اجرا شد. این آزمون نهایی نشان داد که انتخاب ویژگی های مشخص شده توسط PSO توانست عملکرد کلاسیک ها را بهبود ببخشد. و ترکیب الگوریتم PSO برای انتخاب ویژگی ها و SVM برای طبقه بندی، با رسیدن به بالاترین دقت پتانسیل بالایی را برای افزایش دقت و کارایی در طبقه بندی مجموعه داده لنفوگرافی دارا می باشد.

۶- مراجع

1. Oliver, Guillermo and colleagues. "The Lymphatic Vasculature in the 21st Century: Novel Functional Roles in Homeostasis and Disease." 2020.
2. C.M Hassouna and colleagues. "ANN for Evolving Efficient Classification in Lymphography." International Journal of Academic Information Systems Research (IJASIR), 2021.
3. Andrea S. Blevins Primeau. "Data Reveal Growing Burden of Metastatic Cancers in the US", Journal of the National Cancer Institute. September 1, 2022.
4. T. DiSipio and colleagues. "Incidence of unilateral arm lymphoedema after breast cancer: a systematic review and meta-analysis". The Lancet Oncology. 2013.
5. SH. Ridner and colleagues. "A Prospective Study of the Lymphedema and Fibrosis Continuum in Patients with Head and Neck Cancer", LYMPHAT RES BIOL (Lymphatic Research and Biology). 2016.
6. SC. Hayes and colleagues. "Lymphedema following gynecological cancer: Results from a prospective, longitudinal cohort study on prevalence, incidence and risk factors". Gynecologic Oncology. 2017.
7. L. Mango. "Lymphoscintigraphic Indications in the Diagnosis, Management and Prevention of Secondary Lymphedema". Radiation, 2023.
8. P. Sharma and M. Kaur. "Classification in Pattern Recognition: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, April 2013.
9. A. Al Bataineh and S. Manacek, "MLP-PSO Hybrid Algorithm for Heart Disease Prediction", Journal of Personalized Medicine, 2022.

- 10.K. Polat and S. Gunes. "Automated identification of diseases related to lymph system lymphography data using artificial immune recognition system with fuzzy resource allocation mechanism (fuzzy-AIRS)" Biomedical Signal Processing and Control, 2006.
- 11.S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.
- 12.D. M. Mount, S Arya. "ANN: A library for approximate nearest neighbor searching", UMD Computer Science, 2010.
- 13.G.Panchal and coworkers. "Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers" International Journal of Computer Theory and Engineering,2011.
- 14.Abdulaziz Y. Barnawi and Ismail M. Keshta. "Energy Management in Wireless Sensor Networks Based on Naive Bayes, MLP, and SVM Classifications: A Comparative Study", Hindawi Publishing Corporation Journal of Sensors,2016.
- 15.M.Bahoura. "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes", computers in Biology and Medicine,2009.
- 16.M. Zwitter and M. Soklic. the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia
- 17.Chung-Jui Tu and colleagues. "Feature Selection using PSO-SVM", IAENG International Journal of Computer Science, 2006.
- 18.Abdulhamit Subas. "Classification of EMG signals using PSO optimized SVM for diagnosis
- 19.of neuromuscular disorders", Computers in Biology and Medicine, 2013.