



تشخیص سرطان خون نوع لوسمی حاد ALL در تصاویر لام خون با استفاده از روش ترکیبی خوشه‌بندی فازی و شبکه

عصبی

محمد مهدی حسینی\*<sup>(۱)</sup> زهرا سهرابی<sup>(۲)</sup>

(۱) گروه مهندسی کامپیوتر، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

(۲) گروه مهندسی کامپیوتر، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

(تاریخ دریافت: ۱۳۹۹/۶/۱۸ تاریخ پذیرش: ۱۳۹۹/۱۰/۱۷)

چکیده:

تشخیص زودهنگام بیماری سرطان کمک شایانی می‌تواند به روند درمان مبتلایان به این بیماری نماید. بنابراین ارائه روش‌های اتوماتیک جهت تشخیص زود هنگام این بیماری می‌تواند کمک شایانی برای درمان بیماران نماید. در این مقاله یک روش خودکار جهت یادگیری الگوهای تشخیصی پاتولوژیست‌ها، برای تشخیص سلول‌های خونی سرطانی پیشنهاد شده است. روش ارائه شده یک روش ترکیبی کارآمد بر پایه الگوریتم‌های شبکه عصبی و خوشه‌بندی فازی جهت تشخیص سرطان خون نوع ALL است. در این روش ابتدا از الگوریتم FCM جهت بخش‌بندی تصاویر و استخراج ویژگی استفاده شد. سپس به کمک الگوریتم ژنتیک ویژگی‌های مناسب انتخاب شدند و در نهایت جهت تعیین سرطانی بودن تصاویر سلول‌ها با استفاده از شبکه عصبی به دسته بندی تصاویر پرداخته شد. نتایج شبیه‌سازی روش پیشنهادی براساس معیارهای ارزیابی دقت، حساسیت، میانگین هارمونیک دقت - حساسیت و ضریب همبستگی Mcc به ترتیب بطور متوسط برابر ۹۶.۵۳٪، ۸۵.۱۵٪، ۹۵.۲٪ و ۹۰.۸٪ حاصل گردید که حاکی از برتری روش پیشنهادی در تشخیص سرطان، نسبت به دیگر روش‌ها بود.

واژه‌های کلیدی: سرطان خون، خوشه‌بندی فازی، لوسمی لنفوسیتی حاد، دسته‌بندی

\*عهده دار مکاتبات:

محمد مهدی حسینی

نشانی: گروه مهندسی کامپیوتر، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

پست الکترونیک: [hosseini\\_mm@shahroodut.ac.ir](mailto:hosseini_mm@shahroodut.ac.ir) تلفن: ۰۲۳۳۲۳۹۴۲۸۴

## ۱- مقدمه

سرطان‌ها گروهی از بیماری‌ها هستند که به صورت رشد بی‌رویه و خارج از کنترل تعدادی از سلول‌ها ایجاد شده و حدود دویست نوع مختلف دارند [۱]. درمان موفقیت‌آمیز سرطان خون در ده تا پانزده سال اخیر از بزرگ‌ترین دستاوردهای علم پزشکی، به‌خصوص از جنبه سرطان‌شناسی است و در واقع یک انقلاب پزشکی محسوب می‌شود [۲]. به طور معمول سلول‌ها از ساختاری که به آنها داده شده، پیروی می‌کنند و در نتیجه سلامت بدن حفظ می‌شود. اما گاهی این ساختارها در برخی از سلول‌ها درست دریافت نمی‌شود و در نتیجه این سلول‌ها رفتار و واکنشی غیرعادی را در رشد و تکثیرشان پیش می‌گیرند و بدون نظم تکثیر می‌شوند. در بیشتر اقسام، این سلول‌های غیرطبیعی توده‌های توپ‌ر و جامدی را تشکیل می‌دهند که به نام تومور شناخته شده‌اند. با این حال، سلول‌های غیرطبیعی سیستم ایمنی یا خون، به ندرت تومورهای جامد درست می‌کنند. در سرطان خون معمولاً این سلول‌های غیرطبیعی در خون، مغز استخوان و دستگاه لنفاوی به گردش در می‌آید. یکی از انواع سرطان لوسمی است. انواع لوسمی بر اساس سرعت پیشرفت بیماری و بدتر شدن وضعیت بیماری یا براساس نوع گلبول‌های سفید درگیر گروه‌بندی می‌شوند [۲]. این سرطان یکی از انواع سرطان‌های درون مغز استخوان است. واژه لوسمی به معنی خون سفید است و شامل تولید غیرطبیعی سلول‌های خونی سفید می‌شود. تعداد بالای سلول‌های سفید باعث

اختلال توانایی این سلول‌ها در مبارزه با عفونت‌ها شده و همچنین باعث نقص توانایی مغز استخوان برای ساختن سلول‌های قرمز و پلاکت خون می‌شود. لوسمی به دو صورت مزمن یا حاد می‌باشد. سرطان لوسمی مزمن آرام‌تر از نوع حاد آن پیشرفت می‌کند و نوع حاد آن نیاز به درمان سریعتری دارد [۱]. در سرطان خونی نوع حاد، ترکیب خونی بیمار طی ۳-۶ هفته کاملاً به هم خورده و علائم بالینی از جمله افت پلاکت، کاهش گلبول‌های قرمز یا افزایش گلبول‌های سفید، تب و نشانه‌های نارسایی مغز استخوان سریع خود را نشان می‌دهد. سرطان خون حاد دو نوع میلوئیدی<sup>۱</sup> (AML) و لنفوییدی<sup>۲</sup> (ALL) دارد و با توجه به نوع سلول سرطانی، درمان و نوع علائم آنها با هم متفاوت هستند [۲]. منظور از لوسمی لنفوسیتی رشد غیرطبیعی در سلول‌های مغز استخوان که به لنفوسیت تبدیل می‌شوند بوده و نوع میلوئیدی نوعی از سلول‌های مغز استخوان را که تبدیل به سلول‌های خونی از جمله سلول‌های قرمز، سفید و پلاکت می‌شود، تحت تاثیر قرار می‌دهد. برخلاف نوع حاد، سیر بیماری و بروز علائم بالینی در سرطان خون مزمن بسیار کند پیش می‌رود و ممکن است فرد مبتلا به سرطان مدت‌های طولانی از وجود سلول‌های سرطانی در بدنش بی‌اطلاع باشد. سرطان خون مزمن دو نوع میلوپرولیفراتیو<sup>۳</sup> (CML) و لنفوپرولیفراتیو<sup>۴</sup> (CLL) دارد. در بیمارستان‌ها و مراکز درمانی تشخیص سرطان خون با تهیه لام از بافت خون و قراردادن زیر

<sup>1</sup> Myelogenous

<sup>2</sup> Lymphocytic

<sup>3</sup> Myeloproliferative

<sup>4</sup> Lymphoproliferative

داند که روش مورفولوژیکی خودکار برای شناسایی حاد لئوسیتی لوسمی (ALL) توسط تصاویر میکروسکوپ از نمونه خون چگونه است. در این روش ابتدا ارزیابی در مورد شاخص‌های مورفولوژیکی از سلول‌ها انجام شده و در نهایت طبقه بندی برای تشخیص لوسمی صورت می‌پذیرد. حیدری و همکارانش [۴] نمونه‌هایی از ۵۰ بیمار سرطانی (لوسمی حاد لئوسیدی) تهیه و تصاویر مورد نیاز از کلیه سلول‌ها در آزمایشگاه را از لام‌ها استخراج کردند. سپس از این تصاویر به عنوان الگوهای لازم جهت آموزش شبکه‌های عصبی مصنوعی بهره گرفتند. نرم افزاری طراحی شده متشکل از سه شبکه عصبی مجزا و دو واحد تشخیص رنگ و اندازه‌گیری مساحت هسته بود که هر کدام از سه شبکه عصبی بطور جداگانه و با استفاده از تصاویر تهیه شده مورد آموزش قرار می‌گرفتند. این نرم افزار قادر به تشخیص گلبول‌های قرمز و سفید از یکدیگر از لحاظ نرمال یا غیر نرمال بودن بود. نحوه تشخیص رده‌های فوق از یکدیگر بر اساس رنگ هسته، اندازه سیتوپلاسم، اندازه هسته و وجود هستک‌ها درون سیتوپلاسم بود. رافائل مارکوس و همکارانش [۵] یک مطالعه مقایسه‌ای بین انتخاب رو به جلو<sup>۹</sup> (SFS) و الگوریتم ژنتیک<sup>۷</sup> (GA) به عنوان چارچوب کلی برای تجزیه و تحلیل داده‌های بسیار کوچک با هدف شناسایی گروهی از ژن‌ها انجام داده‌اند. برای پیش بینی نتیجه سرطان از شش تکنیک استاندارد و مبتنی بر یادگیری ماشین تجزیه و تحلیل مشخص<sup>۸</sup> (LDA)، ماشین بردار پشتیبان<sup>۹</sup> (SVM)، ناو بیز<sup>۱۰</sup> (NB)، شبکه‌های عصبی سمپتیک، نزدیک‌ترین همسایه<sup>۱۱</sup> (KNN) و پرسپترون چند لایه<sup>۱۲</sup> (MLP) با استفاده از شش مجموعه داده عمومی

میکروسکوپ، توسط یک متخصص پاتولوژی صورت می‌گیرد. پاتولوژیست‌ها با توجه به شکل و تعداد گلبول‌های موجود در خون نوع بیماری را مشخص می‌کنند. تعداد زیاد سلول سفید، یا تعداد بسیار کم سلول قرمز یا پلاکت می‌تواند نشانه ابتلا به سرطان خون باشد. اگر نتیجه ابتلای به لوسمی مثبت باشد پزشکان از یک آزمایش برداشتن بافت استخوان (بافت برداری<sup>۵</sup>) برای مشخص کردن نوع سرطان استفاده می‌کنند. درمان سرطان خون لوسمی بسته به سن، سلامت عمومی و نوع لوسمی متفاوت است. هدف از این تحقیق ارائه یک مدل هوشمند می‌باشد که بتواند با استفاده از شبکه عصبی، الگوهای تشخیصی پاتولوژیست‌ها را آموزش دیده، سپس در تشخیص سلول‌های خونی سرطانی استفاده شود. در روش پیشنهادی نمونه‌های خونی به دو دسته نرمال و غیر نرمال دسته بندی شده که رده غیرنرمال نیز دسته‌بندی لوسمی لئوسیتی حاد (ALL) به سه زیر شاخه مورفولوژیکی L1، L2 و L3 در نظر گرفته شده است. نحوه تشخیص رده‌های بیان شده بر اساس ویژگی‌های هسته شامل رنگ هسته، اندازه هسته، اندازه سیتوپلاسم و وجود هستک‌ها درون سیتوپلاسم می‌باشد.

## ۱- سابقه تحقیق

در تحقیقات انجام شده تاکنون، برخی از آثار با ارزش در مورد سیستم تشخیص خودکار لوسمی با استفاده از تصاویر میکروسکوپی انجام شده است. حلیم و همکارانش [۳] سیستم پیشنهادی با استفاده از ویژگی‌های تصاویر میکروسکوپی به بررسی تغییرات مانند بافت، هندسه، رنگ و تجزیه و تحلیل آماری از تصاویر پیشنهاد کردند. آنها نشان

<sup>9</sup> Support vector machine

<sup>10</sup> Naïve Bayes

<sup>11</sup> K-nearest neighbor

<sup>12</sup> Multiple layer Perceptron

<sup>5</sup> biopsy

<sup>6</sup> Sequential Forward Selection

<sup>7</sup> Genetic Algorithm

<sup>8</sup> Linear Discriminative Analysis

استفاده شده است. نتایج نشان داد پیش بینی نتیجه سرطان با استفاده از GA در مقایسه با یک SFS بهتر جواب می دهد. ملک و همکارانش [۶] روشی را پیشنهاد کردند که بر کاربرد ANN در طبقه بندی نمونه های عادی و غیرعادی خون تمرکز داشت. در این مقاله از ۲۲۰ نمونه خون متشکل از ۶۰ نمونه غیر عادی و ۱۶۰ نمونه سالم در نظر گرفته شد. نتایج به دست آمده نشان داد که طبقه بندی شبکه عصبی می تواند برای این نوع از داده ها نتایج خوبی را نشان دهد. لباتی و همکارانش [۷] تجزیه و تحلیل بصری از نمونه خون peripheral که یک آزمایش مهم در رویه های تشخیص سرطان خون می باشد را مورد استفاده قرار دادند. آنها نشان دادند که سیستم های خودکار بر اساس روش هوش مصنوعی می تواند این عملیات را سرعت بخشیده و همچنین می تواند دقت و صحت پاسخ را در برنامه های کاربردی پزشکی از راه دور افزایش دهد. ولوچامی و همکارانش [۸] روشی مبتنی بر تجزیه و تحلیل تصویر دیجیتال برای طبقه بندی سلول های خونی آسیب دیده و نرمال در تصویر میکروسکوپی ارائه نمودند. آنها تصاویر مربوط به سلول های خون بیماران مختلف را جمع آوری سپس با استفاده عملیات مورفولوژیکی مانند آستانه گیری، تحلیل و مساحت به تقسیم بندی سلول های خون پرداختند. این ویژگی ها از سطح خاکستری و ویژگی های جبری سلول خون استخراج و بر اساس آن پارامترهای هندسی محاسبه شدند. نتایج به دست آمده ثابت کرد که این ویژگی ها بسیار مهم و می توانند برای طبقه بندی استفاده شوند. علاوه بر این، در این پژوهش از شبکه عصبی پس انتشار برای طبقه بندی سلول های خونی استفاده شد. راندمان در طبقه بندی سلول ها طبیعی و غیرطبیعی به ترتیب ۸۰٪ و ۶۶.۶٪ گزارش گردید.

کاکرابورتی و همکارش [۹] مدل های مختلف شبکه عصبی مصنوعی را جهت پیش بینی سطح LDL خون در بیماران مبتلا به دیابت نوع دو بررسی کردند. در این مطالعه از شبکه های عصبی مصنوعی شامل پرسپترون چند لایه و مدل پس انتشار خطا را مورد بررسی قرار گرفت که نتایج حاصل از شبکه های عصبی به کمک یافته های آزمایشگاهی تایید گردید. لوجیا تانگ و همکاران [۱۰] در تحقیق خود یک روش تشخیصی برای آمبولی ریه به کمک شبکه های عصبی مصنوعی و با استفاده از داده های بالینی ایجاد کردند. آنها داده های مربوط به ۱۰۲ بیمار مشکوک به آمبولی ریه را جمع آوری و به بررسی ویژگی های ددیمر، علائم بالینی، اطلاعات دموگرافیک، گازهای خون شریانی پرداختند. در نهایت از یک مدل رگرسیون کمک گرفتند تا ویژگی هایی که همبستگی بیشتری با آمبولی ریه دارند را شناسایی کنند. پس از ورود داده های آموزش و تست شبکه، مدل ایجاد شده نشان داد که هنگامی که مدل رگرسیون و شبکه عصبی پس انتشار خطا در ترکیب با همدیگر مورد استفاده قرار می گیرند با درصد اطمینان بالاتری می تواند کمک به تشخیص بیماری کنند. موهاپاراتا [۱۱] خوشه بندی را برای تقسیم بندی سلول های سفید خون اعمال کرد و برخی ویژگی ها مانند فضا، رنگ، بافت، فراکتال، توصیف فوریه و کانتور استخراج نمود. سیستم طراحی شده برای به رسمیت شناختن لوسمی آموزش دیده بود. قوش [۱۲] یک روش برای پیدا کردن آستانه برای تقسیم بندی لکوسیت معرفی نمودند. اساس روش بر پایه فازی بود و از توابع مختلفی مانند گوسی، گاما، گاجی و غیره در روش ارائه شده استفاده شد. این روش برای بخش بندی هسته به خوبی کار می کند، اما استخراج سیتوپلاسم به عنوان استخراج هسته در تشخیص سرطان، زیاد مورد اهمیت نبود. آنگلو [۱۳] یک سیستم پیشنهاد کرد که در آن الگوریتم دو

مرحله‌ای تقسیم‌بندی خون ارائه شده است. آنها از فیلترهای جفتی و برخی از تکنیک‌های آستانه اتوماتیک استفاده کردند. این سیستم برای استخراج هسته، سیتوپلاسم و هسته از تصاویر لنفوسیت استفاده شد. روند تقسیم‌بندی در دو مرحله است که باعث می‌شود شرایط انتخاب آستانه بهینه برای تقسیم‌بندی دشوار باشد. پوروانتی و همکارانش [۱۴] نشان داده است که بین مدت زمان تغذیه با شیرمادر و ابتلا به لوسمی حاد در کودکان ارتباط معنا دار آماری وجود دارد. نتایج رگرسیون لجستیک نشان داده است که کودکانی که کمتر از یک ماه با شیر مادر تغذیه شده بودند نسبت به کودکانی که ۱۹-۲۴ ماه شیر مادر خورده بودند، ۳/۵۴ برابر شانس بیشتر برای لوسمی حاد داشتند. اوویانگ و همکارانش [۱۵] روشی مبتنی بر شبکه‌های یادگیری عمیق مطرح نمودند. آنها در روش پیشنهادی خود از ۱۳۵۰۴ تصویر بهره گرفتند که ۸۰ درصد از این تصاویر برای آموزش و ۲۰ درصد برای تست مورد استفاده قرار گرفت. شبکه طراحی شده شامل دو قسمت بود که در قسمت اول از تصاویر مجموعه داده MSCOCO برای پیش‌آموزش شبکه استفاده کرده بودند و سپس در ادامه از شبکه CNN برای دسته‌بندی و تشخیص استفاده کردند. نتایج حاصل از روش پیشنهادی ۸۵ درصد بود. کلینتون و همکارانش [۱۶] یک شبکه عصبی با زاویه جداسازی دقیق عمیق (Xception) برای شناسایی سلول‌های leukemic B - lymphoblast ارائه کردند. آنها در روش پیشنهادی خود از الگوریتم‌های شبکه عصبی کانولوشنی مختلف را مورد آزمایش قرار دادند تا بهترین معماری عملکردی برای طبقه‌بندی سلول‌های لوسمی را ارزیابی کنند که منجر به معماری Xception شد. نتایج حاصل از روش پیشنهادی آنها به ترتیب ۹۹٪ و ۹۱٪ دقت در مجموعه آموزشی و تست گزارش گردید. کیلی‌کارسلان و

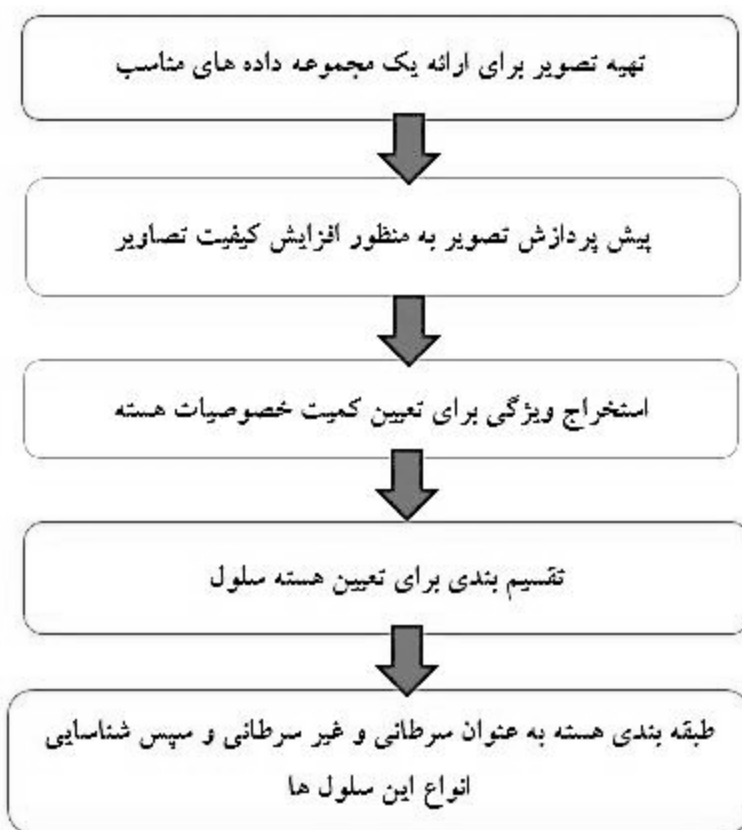
همکارانش [۱۷] روشی با استفاده از ترکیب شبکه یادگیری عمیق CNN و Relif برای تشخیص سرطان پیشنهاد نمودند. آنها در مطالعه خود از سه پایگاه داده آماده OVERIAN، لوسمی و سیستم عصبی بهره گرفتند و با آزمایش شبکه‌های مختلف در نهایت به ترکیب شبکه‌های CNN و Relif دست پیدا نمودند که نتایج حاصل از روش پیشنهادی بر روی هر یک از مجموعه داده‌ها به ترتیب برابر با ۹۸.۶٪، ۹۹.۸۶٪ و ۸۳.۹۵٪ حاصل شد. ساهلول و همکارانش [۱۸] یک رویکرد ترکیبی بهبود یافته برای طبقه‌بندی کارآمد لوسمی ارائه نمودند. آنها ابتدا ویژگی‌هایی را از تصاویر WBC با استفاده از VGGNet استخراج نمودند. سپس مشخصه‌های استخراج شده را با استفاده از الگوریتم ازدحام Salp enhanced enhanced فیلتر کردند. در نهایت روش پیشنهادی خود را بر روی دو مجموعه داده تست و ادعا نمودند که به دقت بالایی دست یافتند و توانسته بودند که پیچیدگی محاسباتی را کاهش دهند.

## ۲- روش پیشنهادی

مطالعات در مقالات پیشین نشان داد که روش‌های ترکیبی معمولاً عملکرد بهتری نسبت به دیگر روش‌ها در تشخیص ارائه نموده‌اند. بنابراین در این مقاله یک روش ترکیبی از سه الگوریتم ارائه شده است. بدین صورت که برای تقسیم‌بندی تصاویر از الگوریتم FCM و برای انتخاب ویژگی از الگوریتم ژنتیک و همچنین از شبکه عصبی برای بهبود تشخیص سلول سرطانی استفاده شده است. شکل ۱ شمای کلی از روش پیشنهادی را نشان می‌دهد. قدم اول برای پیاده‌سازی به دست آوردن تصویر از لام خون و مغز استخوان است. بنابراین برای آماده‌سازی پایگاه داده از این مطالعه، ۲۱ لام خون و مغز

استخوان محیطی از ۱۴ بیمار با ALL و ۷ نفر از افراد سالم استفاده شد. این پایگاه داده بر اساس نمونه‌های موجود در آزمایشگاه پاتولوژی یکی از بیمارستان‌های کشور جمع‌آوری شد. در پردازش تصویر، استخراج ویژگی و حذف نویز از تصاویر، بسیار مهم است. بنابراین یک پیش پردازش اولیه بر روی تصاویر صورت پذیرفت. در این مرحله، ابتدا تصویر ورودی از یک فیلتر میانه برای از بین بردن نویزهای موجود عبور داده می‌شود تا یک تصویر نرم و تمیز بدست آید.

همچنین از این مرحله برای بهبود کیفیت تصویر برای تعیین منطقه مورد نظر، مانند هسته در تصاویر استفاده می‌شود. بعد از انجام مرحله پیش پردازش، تقسیم بندی هسته یکی از کلیدی‌ترین مراحل موجود می‌باشد. عملکرد این مرحله استخراج ویژگی و تقسیم بندی را تحت تاثیر می‌گذارد.



شکل ۱. فلوچارت روش پیشنهادی

## ۱-۲ استخراج ویژگی

در روش پیشنهادی، ویژگی‌های شکل سیتوپلاسم و هسته به عنوان عدد اوایلر، منطقه، پارامتر، قطر، استحکام، محور اصلی،

محور جزئی، خروج از مرکز، جهت گیری، منطقه‌ی محدب، میزان سلول سرطانی و هم چنین سلول‌ها و ویژگی‌های بافت

$$Energy = \sum_{i,j=0}^{n-1} (p_{i,j})^2 \quad (3)$$

$$di = \text{sqrt}(4 \times \frac{\text{area}}{pt}) \quad (4)$$

$$P = 2 \times pi \times r \quad (5)$$

$$\text{area} = \sum_{x=1}^X \sum_{y=1}^Y f(x, y) \quad (6)$$

$$e = \lim_{n \rightarrow \infty} \left[ 1 + \frac{1}{n} \right]^n \quad (7)$$

$$\text{coh} = \sum_i \sum_j \left( \frac{1}{1+(i-f)^2} \right) p(i, j) \quad (8)$$

$$\text{استحکام} = \frac{\text{مساحت}}{\text{مساحت بدنه‌ی محدب}} \quad (9)$$

## ۲-۲ خوشه‌بندی

تفاوت اصلی خوشه‌بندی کلاسیک و خوشه‌بندی فازی در این است که یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. ایده بنیادین در خوشه‌بندی فازی بدین ترتیب است که فرض کنیم هر خوشه مجموعه‌ای از عناصر است. سپس با تغییر در تعریف عضویت عناصر در این مجموعه از حالتی که یک عنصر فقط بتواند عضو یک خوشه باشد (حالت افرازی)، به حالتی که هر عنصر می‌تواند با درجه عضویت - های مختلف داخل چندین خوشه قرار بگیرد، دسته‌بندی‌هایی که انطباق بیشتری با واقعیت دارند، ارائه می‌شوند. در خوشه بندی فازی هر نمونه ورودی می‌تواند متعلق به یک یا چند خوشه باشد و منحصر به یک خوشه نیست. در روش پیشنهادی نیز تعداد خوشه‌ها از قبل مشخص شده است، ۶ دسته برای دسته بندی انتخاب خواهد شد که اعم از: L1, L2,

سالم با GLCM<sup>۱۳</sup> مانند کتراست، همبستگی، انرژی محاسبه شده و آمار همگنی و آنتروپی از ماتریس GLCM هسته و سیتوپلاسم بررسی شده است. مطابق با هماتوپاتولوژیست<sup>۱۴</sup>، کاتور هسته یکی از ویژگی‌های حیاتی برای تمیز دادن لنفوبلاست از لنفوسیت است. برای تجزیه و تحلیل کاتور بخش هسته، بر اساس ویژگی‌ها در منطقه و مرز استخراج شده است. تمام ویژگی‌های استخراج شده از تصویر معادل باینری از بخش هسته با بدون صفر پیکسل نشان دهنده‌ی منطقه‌ی هسته می‌باشد. دو دسته از ویژگی‌های منطقه و مرزی هسته برای ارزیابی کمی استخراج شد. بر اساس مورفولوژی، انواع مختلفی از سرطان خون وجود دارد که شبیه ALL (لوسمی لنفوبلاستی حاد) کوچک، سلول انفجاری لکوسیت یکنواخت و جزء سیتوپلاسم اندک، گرد و معمولاً شامل واحد هستک در داخل هسته است. در حالی که در AML (لوسمی میلوئید حاد<sup>۱۵</sup>)، انفجار، بزرگ‌تر، فرم، نامنظم و معمولاً دارای های متعدد با حضور میله آئور می‌باشد. سلول‌های سفید خون تیره‌تر از پس زمینه به نظر می‌رسد در حالی که سلول - های قرمز خون (گلبول‌های قرمز) در یک سطح شدت متوسط ظاهر می‌شوند. روابط برخی از ویژگی‌ها در معادلات ۱-۹ آورده شده است. در نهایت به منظور دسته بندی داده‌ها در سیستم تشخیصی پیشنهادی از شبکه‌های عصبی مصنوعی استفاده شده است. در این مرحله سلول ها براساس ویژگی - های انتخاب شده در مراحل قبلی به شش گروه اعم از: L1, L2, L3, At, N و Re طبقه بندی می‌گردند.

$$\text{contrast} = \sum_{n=0}^{N_x-1} n^2 \left\{ \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} p(i, j) \right\}, |i-j|=n \quad (1)$$

$$\text{cov} = \frac{\sum_i \sum_j (i, j) p(i, j) - \mu_x \mu_y}{\delta_x \delta_y} \quad (2)$$

15 - Acute myeloid leukaemia

13 Gray-Level co-occurrence matrix

14 - Hematopathologists

$L3, At, N$  و  $Re$  می‌باشد. تابع هدفی که برای این الگوریتم تعریف شده است بصورت معادله ۱۰ می‌باشد.

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2 \quad (10)$$

در فرمول (۱۰)  $m$  یک عدد حقیقی بزرگتر از یک است که در روش پیشنهادی  $m$  برابر ۲ انتخاب شد.  $x_k$  نمونه  $k$ ام و  $v_i$  نماینده یا مرکز خوشه  $i$ ام و  $n$  تعداد نمونه‌ها می‌باشد.  $u_{ik}$  میزان تعلق نمونه  $i$ ام در خوشه  $k$ ام را نشان می‌دهد. علامت  $\|*\|$  میزان تشابه (فاصله) نمونه با (از) مرکز خوشه می‌باشد که می‌توان از هر تابعی که بیانگر تشابه نمونه و مرکز خوشه باشد استفاده کرد. از روی  $u_{ik}$  می‌توان یک ماتریس  $U$  تعریف کرد که دارای  $c$  سطر و  $n$  ستون می‌باشد و مولفه‌های آن هر مقداری بین ۰ تا ۱ را می‌تواند اختیار کنند. اگر تمامی مولفه‌های ماتریس  $U$  بصورت ۰ و ۱ باشند، الگوریتم مشابه  $c$  میانگین کلاسیک خواهد بود. یکی از رایج ترین روش‌های خوشه‌بندی، روش طبقه بندی Fuzzy C-means است. روش FCM مانند روش C-Means جزء خانواده‌ی الگوریتم‌های خوشه بندی دارای تابع هدف قرار دارد که در آنها هدف حداقل کردن یک تابع می‌باشد. الگوریتم FCM با استفاده از اعضای فازی پیکسل‌های مربوط به هر گروه را مشخص می‌کند. این خوشه‌ها به طور مفروض تقریباً هم اندازه هستند. هر خوشه با مرکزش نمایش داده می‌شود. برای محاسبه مرکز خوشه مجموع درجات عضویت هر عنصر به توان  $m$  در خودش به حاصلضرب توان  $m$  درجه عضویت ها تقسیم می‌شود.

## ۲-۳ شبکه عصبی

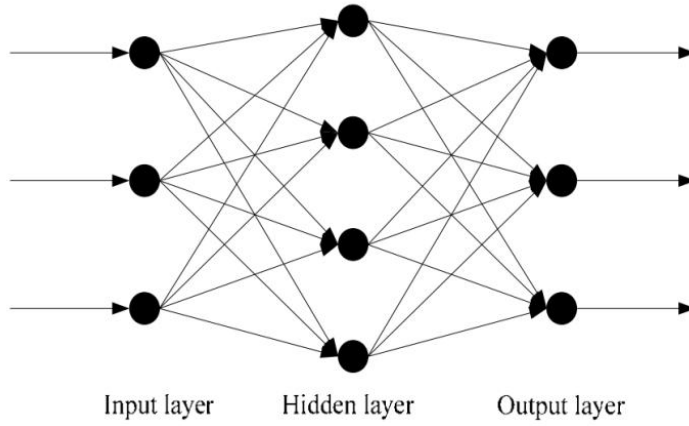
شبکه‌های عصبی مصنوعی در حقیقت به نوعی شبیه‌سازی و مدل‌سازی از سیستم‌های عصبی واقعی‌اند. یکی از ساده ترین

و در عین حال کارآمدترین اتصالات پیشنهادی، مدل پرسپترون چند لایه است. که از یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل یافته است. لازم به ذکر است که تعداد نرون‌های هر لایه مستقل از تعداد نرون‌های لایه‌های دیگر است. جهت آموزش شبکه‌ها روش - های زیادی وجود دارد که یکی از مشهورترین آنها الگوریتم پس انتشار خطا است که از روش‌های با ناظر تلقی می‌شود. واژه‌ی پس انتشار به این معنی است که خطاها به سمت عقب در شبکه تغذیه می‌شوند تا وزن هارا اصلاح کنند. در این الگوریتم ابتدا فرض بر آن است که وزن‌های شبکه به طور تصادفی انتخاب شده اند و در هر مرحله خروجی شبکه محاسبه می‌شود و بر حسب میزان اختلاف آن با خروجی مطلوب، وزن‌ها به گونه‌ای تصحیح می‌شوند که در نهایت خطا به حداقل برسد. خطای کل شبکه به شکل مجموع خطاهای تک تک عصب‌های لایه‌ی خارجی نوشته می‌شود. شکل ۲ ساختار کلی یک شبکه عصبی را نمایش می‌دهد.

## ۳- نتایج آزمایش‌ها و یافته‌ها

در این بخش نتایج حاصل از اجرای الگوریتم تشخیص سرطان لوسمی حاد ارائه شده است. در ابتدا نتایج حاصل از مراحل پیش پردازش اولیه و قطعه بندی صورت گرفته بر روی تصویر نشان داده، سپس به استخراج ویژگی و نحوه آموزش شبکه عصبی پرداخته شده و در نهایت نتایج آورده شده است.

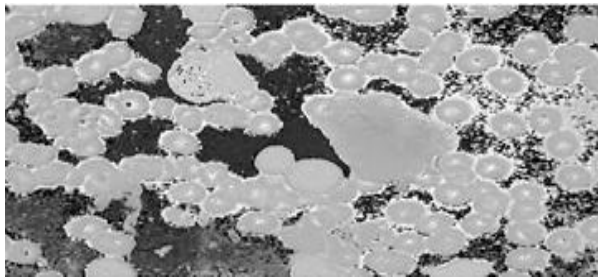




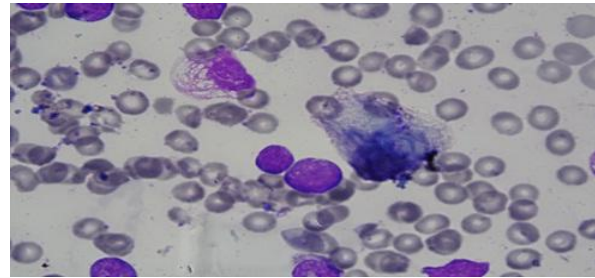
شکل ۲. ساختار کلی مدل شبکه عصبی

(intensity) همان قسمت خاکستری تصویر) است. بر این اساس طبق برنامه مولفه‌ها را جدا کرده و سپس پیش پردازش را روی مولفه V انجام می‌گیرد. شکل ۳ یک نمونه از تصویر اصلی در فضای RGB را به همراه تصاویر حاصل از انتقال به فضای رنگ HSV را نمایش می‌دهد.

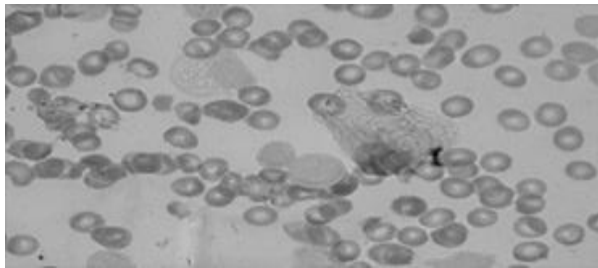
تصاویر موجود در پایگاه داده در فضای رنگ RGB قرار دارند. بنابراین برای افزایش دقت در جداسازی سلول‌های سرطانی، تصاویر به فضای رنگ HSV برده می‌شوند. فضای رنگ HSV فضایی است که در آن H شاخص رنگ، S اشباع رنگ و V مقدار شدت رنگ است. V در فضای رنگی HSV تقریباً برابر



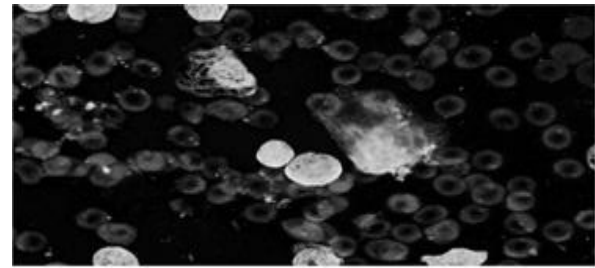
(ب) تصویر H حاصل از فضای HSV



(الف) تصویر اصلی گرفته شده از لام در فضای RGB



(د) تصویر V حاصل از فضای HSV

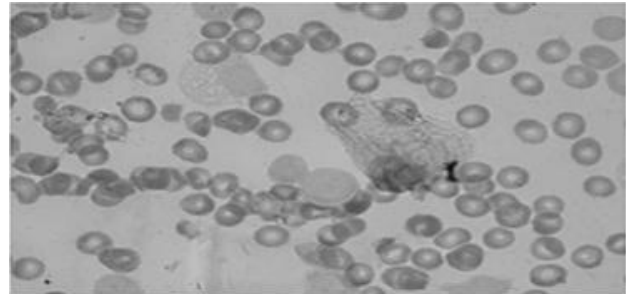
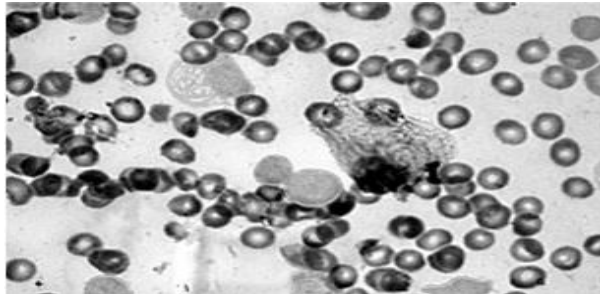


(ج) تصویر S حاصل از فضای HSV

شکل ۳. نمونه تصویر گرفته شده اصلی از روی لام و خروجی حاصل از تبدیل فضای رنگی

موجب کیفیت تصویر شود. شکل ۴ نمونه تصویر قبل و بعد از این نگاشت برای افزایش کنتراست را نشان می‌دهد.

برای انجام پیش پردازش روی  $V$  ابتدا  $1\%$  شدت داده‌های اشباع شده  $V$  را در شدت بالا و پایین نگاشت می‌کنیم. این کار موجب می‌شود که کنتراست تصویر تا حد قابل قبولی افزایش یابد و



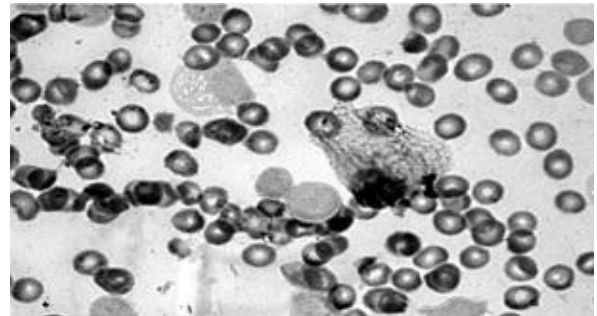
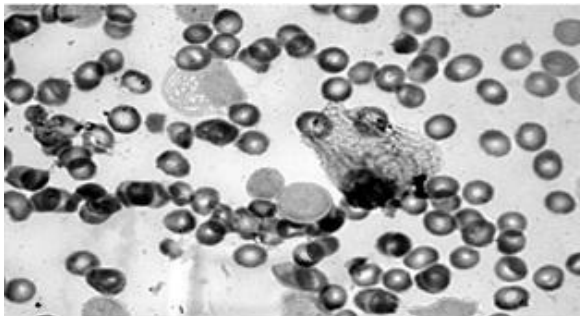
ب) تصویر خروجی حاصل از نگاشت

الف) نمونه تصویر قبل از نگاشت

شکل ۴. تصاویر قبل و بعد از نگاشت اعمال شده در جهت افزایش کنتراست

شکل ۵ مشخص است فیلتر دو بعدی Wiener به خوبی توانسته نویز موجود روی تصویر را حذف کرده و تصویر خروجی مناسبی را در اختیار گذاشته است.

عملیات افزایش کنتراست سبب می‌شود که مقداری نویز به تصویر اضافه شود. از این رو از یک فیلتر Wiener دو بعدی برای از بین بردن این نویز اضافه شده به تصویر استفاده می‌کنیم. نتایج حاصل از بکارگیری فیلترینگ بر روی قسمت  $V$  قبل و بعد از اعمال فیلتر در شکل ۵ نشان داده شده است. همانطور که در



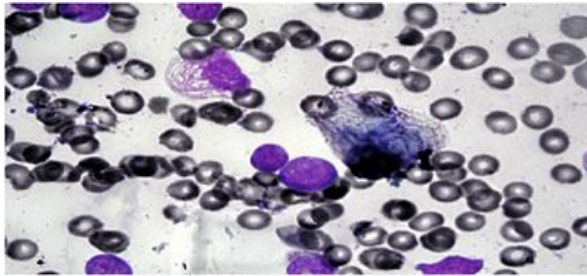
ب) بعد از اعمال

الف) قبل از اعمال

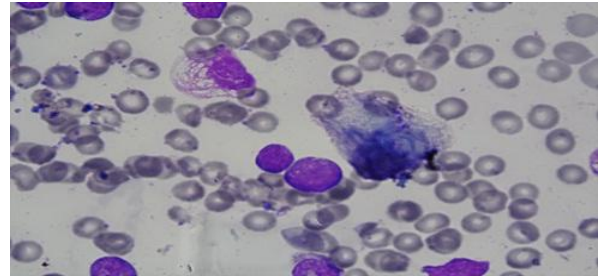
شکل ۵. تصاویر قبل و بعد از اعمال فیلتر wiener

در نهایت این تصویر V بهینه شده با دو بخش H و S اولیه ترکیب شده و تصویر خروجی جدید RGB از این مولفه ها بازسازی می شود.

شکل ۶ تصویر خام اولیه و تصویر نهایی پیش پردازش شده را نشان می دهد.



(ب)

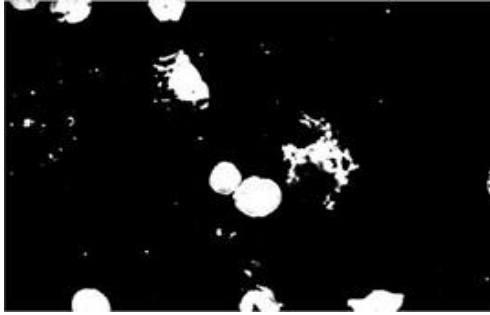


(الف)

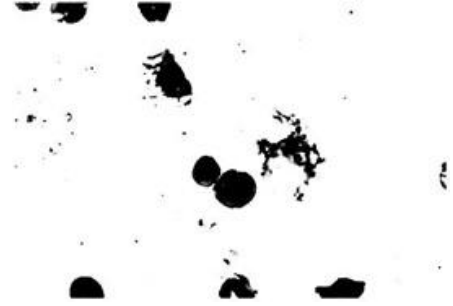
شکل ۶. الف) نمونه تصویر اصلی اولیه ب) نمونه تصویر حاصل از اعمال پیش پردازش

همانطور که در شکل ۸ مشخص شده است. به طور تجربی می توان حدس زد که تصویری که مرکز آن نزدیک تر به مقدار بازه ۸۰ تا ۱۰۰ است به عنوان دسته nuclei انتخاب می شود. پس از قطعه بندی و جداسازی nucleus در تصویر، ویژگی های GLCM استخراج شده و تعداد ۱۰ ویژگی اصلی و تاثیرگذار استخراج می شود. بعد از ایجاد ماتریس ویژگی از روی تصاویر موجود برای تشخیص نمونه های سرطانی از شبکه عصبی استفاده گردید. برای انتخاب بهترین شبکه، نمونه های مختلفی از شبکه ها پیاده سازی و مورد مورد بررسی قرار گرفت. جدول ۱ برخی از نتایج حاصل از بکارگیری انواع مختلف شبکه های عصبی را نشان می دهد.

پس از پیش پردازش اولیه، از الگوریتم Kmeans برای قطعه بندی استفاده نمودیم. الگوریتم Kmeans تصویر رنگی را در چهار برچسب رنگی متفاوت قطعه بندی می کند. سپس کمینه میانگین این چهار برچسب رنگی به عنوان تصویر مطلوب انتخاب شدند. این تصویر مجدداً وارد الگوریتم قطعه بندی Fuzzy-Cmeans می شود. در این مرحله تصویر را به دو قطعه مختلف تقسیم بندی شده که می توان از این تصاویر برای استخراج ویژگی استفاده نمود. تصویر خروجی حاصل از روش FCM در شکل ۷ و هیستوگرام مربوط به قبل و بعد از بخش بندی تصویر در شکل ۸ نشان داده شده است.

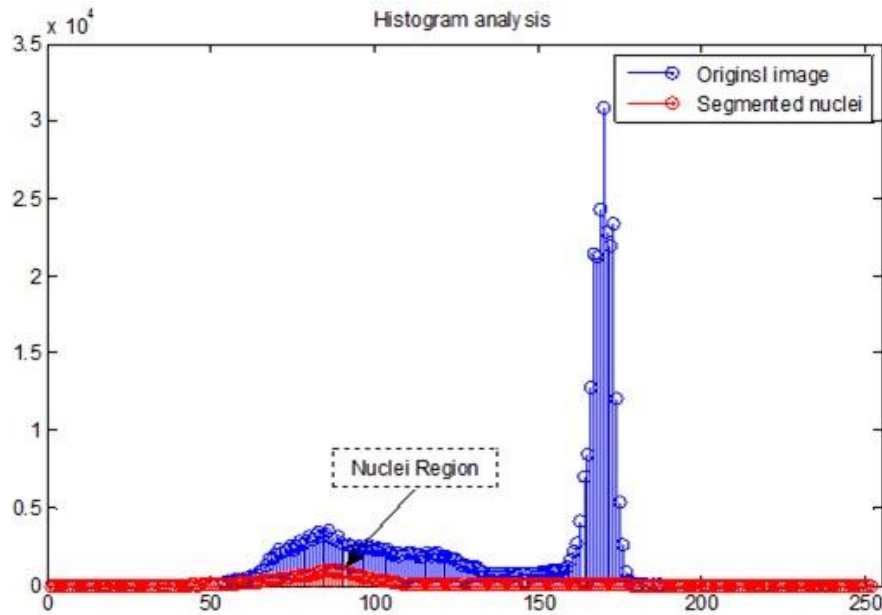


ب



الف

شکل ۷. تصاویر حاصل از اعمال الگوریتم FCM



شکل ۸. نمودار هیستوگرام تصویر اصلی و تصویر قطعه‌بندی شده

شد. لازم به ذکر است که امکان بالابردن تعداد تکرارها برای رسیدن به دقت بالاتر وجود داشت اما پس از چند آزمایش معلوم شد که بالابردن تعداد تکرارها تنها باعث ایجاد پدیده "پس تخمین"<sup>۲۰</sup> شده و مانع از عملکرد مناسب شبکه در تست می‌شود. دقت تست شبکه در این حالت ۹۲.۶۶٪ است. سپس شبکه PNN<sup>۲۱</sup> را با استفاده از داده‌های استخراج شده آموزش داده شد. خروجی تست این شبکه نیز دقتی در حدود

دقت در شبکه‌های MLP<sup>۱۶</sup> و پیش‌خور<sup>۱۷</sup> با توابع آموزش مختلف و تعداد لایه‌ها و نرون‌های متفاوت کمتر از ۵۰٪ حاصل شد. بهترین جواب در این حالت در شبکه MLP با استفاده از تابع انتقال HARDLIM و تابع آموزش LEARNP و تعداد لایه‌های پنهان ۲۵ آموزش داده شد که مقدار ۴۸٪ برای تست آن بدست آمد. شبکه RBF<sup>۱۸</sup> سپس مورد استفاده قرار گرفت. در این شبکه تعداد ۵۱۵ تکرار<sup>۱۹</sup> در نظر گرفته

<sup>19</sup> epoch

<sup>20</sup> Overestimation

<sup>21</sup> Probabilistic Neural Network

<sup>16</sup> MultiLayer Perceptron

<sup>17</sup> FeedForward

<sup>18</sup> Radial Basis Function

داده به ترتیب برابر ۱۲ و ۸ حاصل گردید. در این شبکه می-توان با استفاده از پارامتر پهنا<sup>۲۲</sup> میزان صافی<sup>۲۳</sup> تخمین‌گر را تنظیم نمود به گونه‌ای که با کاهش مقدار پهنا داده‌ها به تخمین‌گر نزدیک‌تر شده و بالعکس با افزایش آن میزان صافی افزایش می‌یابد. شکل ۹ میزان دقت شبکه بر اساس مقدار پهنا را نشان می‌دهد. در شبکه طراحی شده برای لایه‌های پنهان از توابع اشباع *logsig* استفاده گردید. تعداد *epoch*ها برابر ۱۰۰۰ در نظر گرفته شده است.

۹۲.۶٪ داشته اما دقت آموزش آن از شبکه RBF بالاتر بدست آمد. آزمایش‌ها بر روی داده‌ها نشان داد که شبکه GRNN زیر بالاترین دقت را در مقایسه با سایر شبکه‌های بیان شده دارد. این شبکه به گونه‌ای شبیه به RBF است. شبکه طراحی شده شامل چهار لایه می‌باشد. که از این چهار لایه، دو لایه آن نهان و دو لایه دیگر ورودی و خروجی می‌باشند. تعداد نودهای ورودی ۱۰ و تعداد نودهای خروجی ۶ در نظر گرفته شد. تعداد نودهای لایه نهان متناسب با نمونه تصاویر پایگاه

جدول ۱ نتایج حاصل از شبکه‌های عصبی مختلف

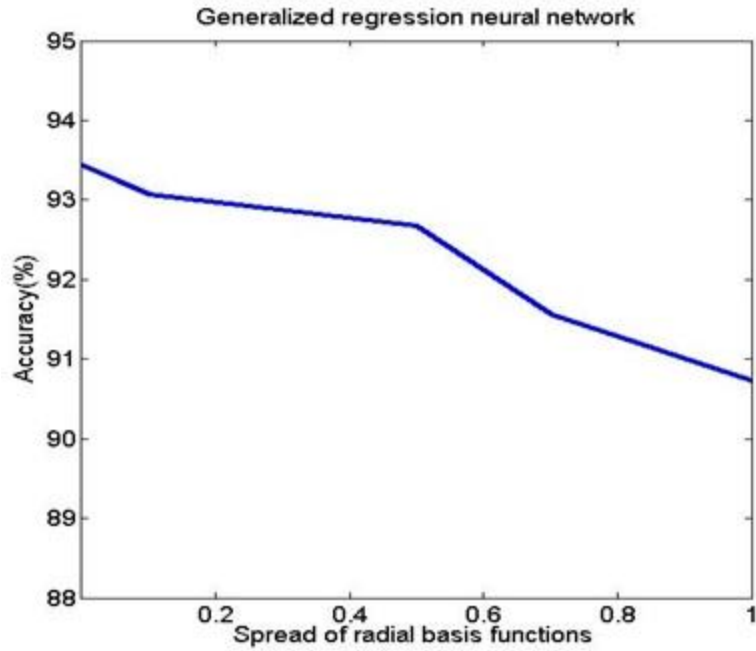
Elapsed time(sec)	MSE	مدل شبکه عصبی
12.7	1.056e-01	Resilient Backpropagation
18.3	8.412e-01	BFGS quasi-Newton backpropagation
4.7	5.214e-01	Gradient descent with momentum and adaptive learning rate backpropagation
89.3	2.626e-01	Levenberg-Marquardt backpropagation
3.6	7.36e-3	Generalized regression neural network
3.78	9.83e-3	Probabilistic neural network

$$\text{Sensitivity} = \frac{T_p}{T_p + F_n} \quad (11)$$

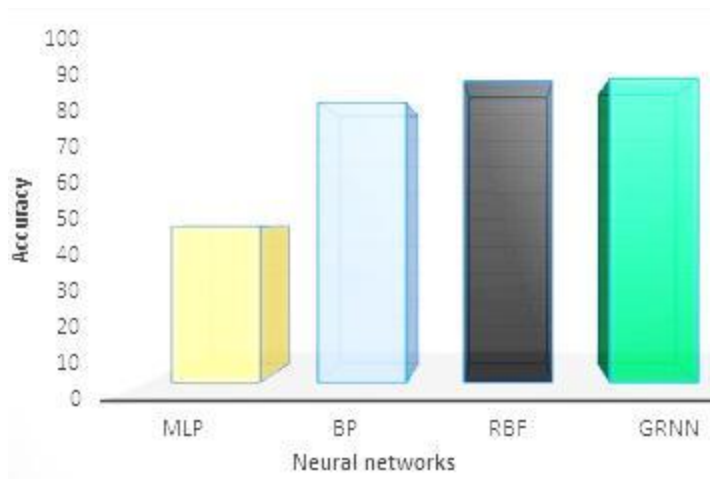
$$\text{Specificity} = \frac{T_n}{T_n + F_p} \quad (12)$$

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (13)$$

بالاترین دقت دسته‌بندی تست این شبکه مقدار ۹۳.۴۴٪ بدست آمده در این شبکه به ازای  $\text{Spread} < 1e-2$  است. از طرفی دیگر سرعت آموزش و ساخت این شبکه به مراتب بالاتر از شبکه‌های دیگر است. شکل ۱۰ میزان دقت شبکه‌های مختلف در تشخیص سرطان ALL را نمایش می‌دهد. معیارهای ارزیابی روش پیشنهادی در معادلات ۱۱ تا ۱۳ آورده شده است.



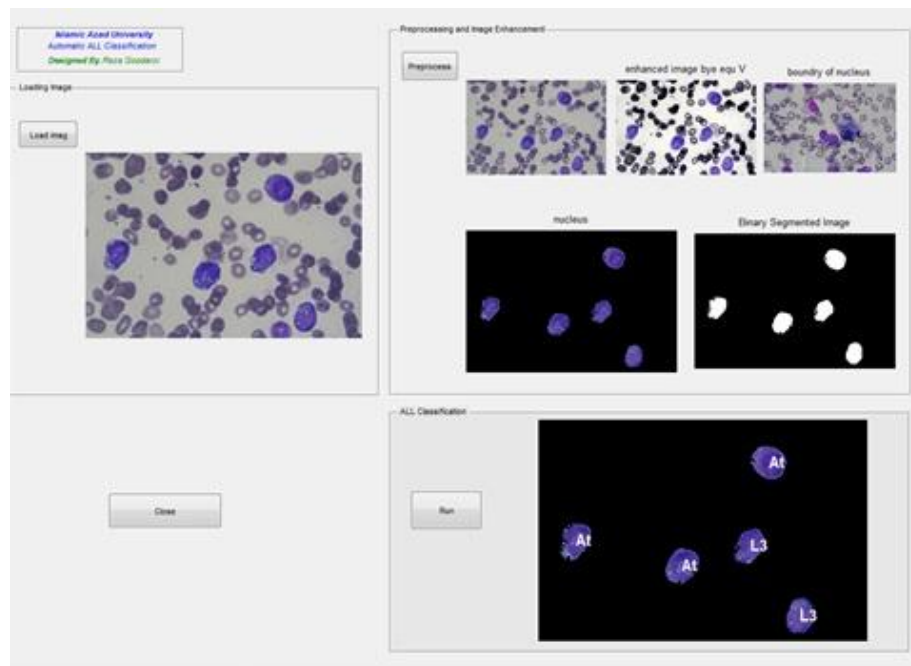
شکل ۹. نمودار تغییرات دقت به تغییرات Spread در شبکه GRNN



شکل ۱۰. مقایسه نتایج حاصل از شبکه‌های مختلف

دقت و حساسیت (بطور متوسط ۹۵.۲٪) و ضریب همبستگی MCC (بطور متوسط ۹۰.۸٪) گزارش گردید. نتایج حاصل از شبیه سازی برتری روش پیشنهادی را از لحاظ دقت، ویژگی، مقدار پیش فرض مثبت، مقدار پیش فرض منفی، میانگین هارمونیک دقت و حساسیت و ضریب همبستگی MCC در تشخیص سرطان، نسبت به روش‌های مشابه دیگر نشان داد. شکل ۱۱ شمایی از رابط کاربری طراحی شده برای استفاده از مدل پیشنهادی توسط پزشکان را نمایش می‌دهد.

که TP تعداد تصاویر تومور که به درستی سرطان تشخیص داده می‌شوند، FP تعداد تصاویر غیر سرطانی که به اشتباه سرطان تشخیص داده می‌شوند، TN تعداد تصاویر غیر سرطانی که به درستی سرطان تشخیص داده نمی‌شوند، FN تعداد تصاویر سرطان که به اشتباه غیر سرطانی تشخیص داده می‌شوند. در نهایت با توجه به معیارهای ارزیابی در نظر گرفته شده دقتی برابر با ۹۶/۵۳ درصد حاصل گردید. همچنین بر اساس دیگر معیارهای ارزیابی، حساسیت بالا (بطور متوسط ۸۵.۱۵٪)، ویژگی (بطور متوسط ۹۸.۱۷)، مقدار پیش فرض مثبت (بطور متوسط ۹۵.۲٪)، مقدار پیش فرض منفی (بطور متوسط ۹۵.۶٪)، میانگین هارمونیک



شکل ۱۱. شمایی از رابط کاربری طراحی شده

دارند و بدون داده کافی نتایج مناسبی از خود نشان نمی‌دهند. هنوز میتوان روی روش‌های کلاسیک کار نمود. با بررسی نتایج مقالات متعدد نشان می‌دهد که روش‌های معرفی شده، به ندرت کارایی بالایی دارند و این به این معناست که هنوز تعداد زیادی از نتایج تشخیص سرطان خون، نادرست عمل می‌کنند. یکی از دلایل این اتفاق، عدم انتخاب ویژگی مناسب است که سبب می‌شود خروجی الگوریتم‌های کلاس بندی

#### ۴- نتیجه گیری

در سال‌های اخیر در الگوریتم‌های تشخیص سرطان خون پیشرفت‌های زیادی صورت گرفته، اما این پیشرفت‌ها هنوز کافی نیست. هرچند امروزه تکنیک یادگیری عمیق به عنوان یک روش کارآمد در این حوزه معرفی شده است. ولی با توجه به اینکه اینگونه روش‌ها به حجم بالایی از داده نیاز

صحت صد درصدی در تشخیص نداشته باشند. بنابراین در روش پیشنهادی سعی گردید که با معرفی ویژگی مناسب این معضل رفع گردد. همچنین در روش پیشنهادی علاوه بر انتخاب الگوریتم کلاس بندی مناسب، انتخاب ویژگی های متناسب با الگوریتم ژنتیک، تاثیر زیادی در درستی روش های تشخیصی ضایعات از خود نشان داد. این تحقیق گامی است کوچک در جهت طراحی و ساخت سیستمی جامع با قدرت تشخیص انواع سرطان های خون که می تواند در مراکز درمانی و در نقاط دور دستی که دسترسی به متخصص پاتولوژی وجود ندارد توسط پزشکان و متخصصان دیگر با در اختیار داشتن یک دوربین مخصوص تصاویر پزشکی و یک دستگاه کامپیوتر مورد استفاده قرار گیرد. نتایج حاصل از روش پیشنهادی حاکی از بهبود دقت بدست آمده نسبت به روش های پیشین است. در نهایت پیشنهاد می گردد با جمع آوری داده های مناسب از تکنیک های موجود در یادگیری عمیق بهره برد.



[1] Abdeldaim, Ahmed M., Ahmed T. Sahlol, Mohamed Elhoseny, and Aboul Ella Hassanien. "Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis." In *Advances in Soft Computing and Machine Learning in Image Processing*, pp. 131-147. Springer, Cham, 2018.

[2] Liu, Ying, and Feixiao Long. "Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning." In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, pp. 113-121. Springer, Singapore, 2019.

[3] Abd Halim, Nurul Hazwani, Mohd Yusoff Mashor, and Rosline Hassan. "Automatic blasts counting for acute leukemia based on blood samples." *International Journal of Research and Reviews in Computer Science* 2, no. 4 2011.

[۴] حبیب حیدری، امیرابوالفضل صورتگر، ایران رشیدی، نغمه مالکپور، ابوذر پرویزی، "تشخیص سرطان خون نوع خاصی از لوسمی حاد لنفوبلاستیک با استفاده از شبکه های عصبی مصنوعی"، سیزدهمین کنفرانس دانشجویی مهندسی برق ایران دانشگاه تربیت مدرس ۱۳۹۸.

[5] Luque-Baena, Rafael Marcos, Daniel Urda, Jose Luis Subirats, Leonardo Franco, and Jose M. Jerez. "Application of genetic algorithms and constructive neural networks for the analysis of microarray cancer data." *Theoretical Biology and Medical Modelling* 11, no. S1 2014.

[6] Adjouadi, Malek, Melvin Ayala, Mercedes Cabrerizo, Nuannuan Zong, Gabriel Lizarraga, and Mark Rossman. "Classification of leukemia blood samples using neural networks." *Annals of biomedical engineering* 38, no. 4, PP:1473-1482, 2010.

[7] Labati, Ruggero Donida, Vincenzo Piuri, and Fabio Scotti. "All-IDB: The acute lymphoblastic leukemia image database for image processing." In *2011 18th IEEE International Conference on Image Processing*, pp. 2045-2048. IEEE, 2011.

[8] Veluchamy, Magudeeswaran, Karthikeyan Perumal, and Thirumurugan Ponuchamy. "Feature extraction and classification of blood cells using artificial neural network." *American journal of applied sciences* 9, no. 5, 2012.

[9] Chakraborty, Monisha, and Bipan Tudu. "Comparison of ANN models to predict LDL level in Diabetes Mellitus type 2." In *2010 International Conference on Systems in Medicine and Biology*, pp. 392-396. IEEE, 2010.

[10] Tang, Lujia, Lina Wang, Shuming Pan, Yi Su, and Ying Chen. "A neural network to pulmonary embolism aided diagnosis with a feature selection approach." In *2010 3rd International Conference on Biomedical Engineering and Informatics*, vol. 6, pp. 2255-2260. IEEE, 2010.

[11] Saritas, Ismail, Ilker Ali Ozkan, and Ibrahim Unal Sert. "Prognosis of prostate cancer by artificial neural networks." *Expert Systems with Applications* 37, no. 9, 6646-6650, 2010.

[12] Kline, Jeffrey A., Dawn Neumann, Cassandra L. Hall, and Jacob Capito. "Role of physician perception of patient smile on pretest probability assessment for acute pulmonary embolism." *Emergency Medicine Journal* 34, no. 2, PP: 82-88, 2017.

[13] Jeong, Young-Seon, Kyoung Seok Shin, and Myong K. Jeong. "An evolutionary algorithm with the partial sequential forward floating search mutation for large-scale feature selection problems." *Journal of The Operational research society* 66, no. 4, PP:529-538, 2015.

[14] Purwanti, Endah, and Evelyn Calista. "Detection of acute lymphocyte leukemia using k-nearest neighbor algorithm based on shape and histogram features." In *Journal of Physics: Conference Series*, vol. 853, no. 1, p. 012011. IOP Publishing, 2017.

[15] Ouyang, Nengliang, Weijia Wang, Li Ma, Yanfang Wang, Qingwu Chen, Shanhong Yang, Jinye Xie et al. "Diagnosing acute promyelocytic leukemia by using convolutional neural network." *Clinica Chimica Acta* 512, 2020.

[16] Clinton Jr, Laurence P., Karen M. Somes, Yongjun Chu, and Faizan Javed. "Acute Lymphoblastic Leukemia Detection Using Depthwise Separable Convolutional Neural Networks." *SMU Data Science Review* 3, no. 2, 2020.

[17] Kilcarslan, Serhat, Kemal Adem, and Mete Celik. "Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network." *Medical Hypotheses*, 2020.

[18] Sahlol, Ahmed T., Philip Kollmannsberger, and Ahmed A. Ewees. "Efficient classification of white blood cell leukemia with improved Swarm optimization of deep features." *Scientific Reports* 10, no. 1, 20