الگوریتم بهینهٔ تقسیم‌بندی تصاویر میکروسکوپی خون برای تشخیص سلول‌های لوسمی حاد لنفوبلاست با استفاده از الگوریتم FCM و بهینه‌سازی ژنتیک

عباس کریمی* و لیلاسادات حسینی

گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی اراک، اراک، ایران

چکیده

سرطان‌ها گروهی از بیماری‌ها هستند که به‌صورت رشد بی رویه و خارج از کنترل تعدادی از سلول‌ها، ایجاد می‌شوند و حدود دویست نوع مختلف دارند. سرطان لوسمی (خون) یکی از انواع این سرطان‌هاست. تشخیص سرطان خون در بیمارستان‌ها و مراکز درمانی با تهیه لام از بافت خون و قراردادن زیر میکروسکوپ و توسط یک متخصص پاتولوژی صورت می‌گیرد. پاتولوژیست‌ها با توجه به شکل و تعداد گلبول‌های موجود در خون نوع بیماری را مشخص می‌کنند. هدف از این مقاله ارائه مدلی هوشمند با استفاده از الگوریتم FCM [1] به‌منظور خوشه‌بندی و شبکهٔ عصبی برای انتخاب ویژگی‌هاست؛ همچنین در آن از الگوریتم ژنتیک در مرحله بهبود الگوهای تشخیصی استفاده شده است. با استفاده از این مدل به تشخیص زودهنگام سرطان لوسمی حاد لنفوبلاست و سپس دسته‌بندی ALL[2] به سه زیر شاخه مورفولوژیکی (L1، L2 و L3) می‌توان اقدام کرد. در این پژوهش نمونه‌هایی از ۳۸ بیمار سرطانی لوسمی حاد لنفوبیدی تهیه شد. این مطالعه بر روی ۶۸ تصویر میکروسکوپی و با درنظرگرفتن پانزده ویژگی هندسی و آماری انجام شد که نتیجه آن حاکی از حساسیت، ویژگی و دقت بالاتر برای ده ویژگی نسبت به سایر ویژگی‌ها بود. بر اساس ویژگی‌های استخراج‌شده، این روش با سه روش مشابه اخیر مقایسه شد. ارزیابی‌ها نشان داد که روش پیشنهادی به‌طور میانگین پارامترهای حساسیت، ویژگی و دقت را به میزان ۸۵/۱۵٪، ۹۸/۱۷٪ و ۹۶/۵۳٪ به‌دست آورد.

واژگان کلیدی: سرطان خون، الگوریتم FCM، شبکهٔ عصبی، الگوریتم ژنتیک، خوشه‌بندی

An Optimal Algorithm for Dividing Microscopic Images of Blood for the Diagnosis of Acute Pulmonary Lymphoblastic Cell Using the FCM Algorithm and Genetic Optimization

Abbas Karimi* & Leila Sadat Hosseini

Department of Computer Engineering, Faculty of Engineering, Arak Branch, Islamic Azad University, Arak, Iran

Abstract

Cancer is type of disease caused by irregular, uncontrollable growth of blood cells in bone marrow. The process of generating three main blood cells including pallets, red and white blood cells, is started from a progenitor cell called as blast. Blast generates a considerable number of immature cells which are developed affected by differentiation factors. If any interruption occurs during this process, leukemia may be initiated. Diagnosis of leukemia is performed at hospitals or medical centers by examination of the blood tissue smeared across a slide and under a microscope by a pathologist. Processing the digital images of blood cells, in order

[1] Fuzzy C-means Clustering

[2] Acute Lymphoblastic Leukemia

* Corresponding author

* نویسندهٔ عهده‌دار مکاتبات

to improve the quality of the image or highlighting the malicious segments of the image, is important in early stages of the disease.

There are four types of leukemia consisting acute or chronic and myeloid or lymphocytic. Acute lymphocytic (or lymphoblastic) leukemia (ALL) is concentrated in this study. ALL is caused by continuous generation of immature, malignant lymphocytes in bone marrow which are speeded by blood circulation to other organs.

In this research, fuzzy C-means (FCM) algorithm is applied to blood digital images for clustering purpose, neural networks for feature selection and Genetic Algorithm (GA) for optimization. This model diagnoses ALL at early stages and categorizes it into three morphological subcategories (i.e., L1, L2, and L3). For performance evaluation of the proposed method, 38 samples of patients with ALL were collected. It was performed on 68 microscopic images in terms of 15 features and yielded to higher percentage of sensitivity, specificity, and accuracy for 10 out of 15 features. The proposed method was compared to three recent methods. The evaluations showed that the sensitivity, specificity and accuracy reached to 85.15%, 98.17% and 96.53%, respectively.

Keywords: leukemia, FCM algorithm, neural network, genetic algorithm, clustering

۱- مقدمه

درمان موفقیت‌آمیز سرطان خون در سال‌های اخیر از بزرگ‌ترین دستاوردهای علم پزشکی، به‌خصوص از جنبهٔ سرطان‌شناسی است و درواقع یک انقلاب پزشکی محسوب می‌شود که تا کنون جان هزاران تن را نجات داده است [1].

سلول‌های خون به سه دسته گلبول‌های سفید، گلبول‌های قرمز و پلاکت‌ها و گلبول‌های سفید خود به دو دسته میلوییدها و لنفوسیت‌ها تقسیم‌بندی می‌شوند. لنفوسیت‌ها نیز به دو دسته نرمال و غیر نرمال دسته‌بندی شده و همچنین گلبول‌های سفید نیز در صورت نرمال یا غیر نرمال‌بودن از هم تشخیص داده شده که رده غیر نرمال خود به سه بخش L1، L2 و L3 تقسیم‌بندی می‌شود که حتی نوع آنها هم از یکدیگر قابل تشخیص است. نحوه تشخیص رده‌های فوق از یکدیگر بر اساس رنگ هسته، اندازه سیتوپلاسم، اندازه هسته و وجود هستک‌ها درون سیتوپلاسم است.

سرطان خون در اثر افزایش تعداد سلول‌های غیرطبیعی و به‌اصطلاح سلول‌های بلاست در مغز استخوان اتفاق می‌افتد. مغز استخوان برای اینکه بتواند سلول‌های طبیعی شامل گلبول‌های قرمز، گلبول‌های سفید و پلاکت را بسازد، از یک سلول اجدادی به نام «بلاست» آغاز به کار می‌کند. این سلول قدرت زایش و تکثیر بالایی دارد و به‌سرعت تعداد زیادی سلول نارس می‌سازد. این سلول‌های نارس تحت تأثیر مواد رشدی، کامل شده و به سلول‌های تخصص یافته خونی تبدیل می‌شوند. اگر در روند رشد سلول‌های نارس وقفه ایجاد شود، سرطان خون پدید می‌آید [2].

لوکمی لنفوبلاستیک حاد (ALL) نوعی از لوکمی یا سرطان گلبول‌های سفید خون است که با افزایش تعداد لنفوبلاست‌ها نمود می‌کند. در این بیماری گلبول‌های سفید بدخیم و غیربالغ بی‌وقفه در مغز استخوان تکثیر شده و با ممانعت از تولید سلول‌های سالم و بالغ و انتشار به سایر ارگان‌ها سبب بیماری و مرگ می‌شوند. بیشترین شیوع لوکمی لنفوبلاستیک حاد در کودکان ۲ تا ۵ سال دیده می‌شود. از تعداد کل مبتلایان حدود ۸۰٪ کودکان درمان شده و در حدود ۶۰-۴۵٪ بالغین سال‌های طولانی از بیماری رها خواهند بود [2].

تشخیص این بیماری با شرح حال، معاینهٔ بالینی، شمارش کامل سلول‌های خونی و تهیه اسمیر مغز استخوان صورت می‌گیرد. با توجه به بروز علائم عمومی در این بیماری، بررسی و رد سایر علل احتمالی بروز ناخوشی در کودکان لازم است. درکل هر چه تعداد گلبول‌های سفید خون بیشتر باشد، پیش‌آگهی بدتری صورت می‌گیرد [3].

بیماری سرطان خون با توجه به ویژگی‌های سلول‌های بدخیم و نابالغ تولیدشده به سه دسته ALL 1,2,3 تقسیم‌بندی می‌شود [4].

شیمی‌درمانی خط نخست درمان این بیماری و شامل سه فاز Remission Induction، Intensification و Maintenance Therapy است. در مرحلهٔ نخست فائق آمدن بر لوسمی (القائی) (Remission Induction) با هدف کشتن سریع بیشینهٔ سلول‌های توموری صورت می‌گیرد؛ به‌طوری‌که کمتر از ۵٪ بلاست‌ها در مغز استخوان باقی مانده، سلول‌های خونی طبیعی و سلول‌های توموری از خون حذف می‌شود. در این مرحله از ترکیب پردنیزولون یا دگزامتازون، وینکریستین و آسپارژیناز استفاده می‌شود.

در مرحلهٔ دوم درمان، جهت جلوگیری از عود (تثبیت) (Intensification) هدف نابودکردن هر چه بیشتر سلول‌های

توموری است. در بسیاری از پروتکل‌های درمانی استفاده از پروفیلاکسی و درمان درگیری CNS در مرحلهٔ نخست و دوم مدنظر است. در این مرحله نیز از انواع داروهای شیمی‌درمانی همراه با کورتیکواستروئیدها استفاده می‌شود.

در مرحلهٔ سوم پیش‌گیری از عود در دراز مدت (نگهدارنده) (Maintenance Therapy) با هدف نابودی همه سلول‌های توموری انجام می‌شود؛ به‌طوری‌که هیچ سلول توموری باقی نماند. چون حتی در صورت باقی‌ماندن تعداد اندکی از آنها، بیماری دوباره عود خواهد کرد. در این مرحله نیز داروهای شیمی درمانی همراه با کورتیکواستروئیدها تجویز می‌شوند. این مرحله برای پسران سه سال و برای دختران و بالغین دو سال طول می‌کشد [5].

روش‌های دیگر درمانی چون رادیوتراپی، درمان‌های بیولوژیک [6] و ایمونوتراپی نیز می‌تواند مورد استفاده قرار گیرد. با توجه به افزایش زمان و هزینه درمان در این بیماری، ارایهٔ روش‌هایی که با سرعت و دقت قابل قبولی بتوانند به پیش‌بینی و یا تشخیص سرطان کمک کنند همواره مورد توجه است. تا کنون روش‌های مختلفی برای تشخیص زودهنگام سرطان ارائه شده است که هر کدام به نوبهٔ خود قوت و ضعف‌هایی دارند.

از میان روش‌های مختلف تشخیص سرطان، روش‌های مبتنی بر پردازش تصویر جایگاه ویژه‌ای دارند. در تمام این روش‌ها استفاده از روش‌های پردازش تصاویر دیجیتال به شکل چشم‌گیری به افزایش کیفیت تصویر بافت و درنتیجه افـزایش دقـت و تشخیص صحیح ضایعهٔ سـرطانی منجر می‌شود.

۲- بررسی متون

در پژوهشی [7] تصاویر میکروسکوپی به‌دست‌آمده از خون و مغز استخوان اسمیر بیماران مبتلا به لوسمی حاد لنفوبلاستیک و افراد نرمال پس از اعمال پردازش تصویر، هسته سلول‌ها توسط الگوریتم k-means تقسیم‌بندی شد؛ سپس ویژگی‌های هندسی و آماری از هسته، استخراج و درنهایت این سلول‌ها به سلول‌های سرطانی و غیر سرطانی با استفاده از بردار پشتیبانی ماشین ACM طبقه‌بندی شدند. این سلول‌ها نیز خود به چند نوع توسط ماشین بردار پشتیبان طبقه‌بندی شدند. حساسیت، ارزش و دقت برای سلول‌های سرطانی و غیر سرطانی ۹۸٪، ۹۵٪ و ۹۷٪ بود. طبقه‌بندی توسط این پارامترها انجام و مورد بررسی قرار گرفت. این پارامترها نیز برای ارزیابی سلول به‌طور متوسط ۸۴/۳٪، ۹۷/۳٪ و ۹۵/۶٪ بود.

در پژوهشی دیگر [8] با استفاده از خصوصیت متمایز هسته از بقیه اجزای یک لام خونی، ابتدا آن‌ها را با استفاده از الگوریتمی مبتنی بر روش خوشه‌بندی k-means تشخیص داده، سپس به جداسازی سیتوپلاسم از بقیه تصویر لام با تکیه بر حد آستانه بر مبنای هیستوگرام سطح خاکستری پرداخته شد. الگوریتم مذکور جهت استخراج هسته و سیتوپلاسم بر روی یکصد دادهٔ میکروسکوپی اعمال شد که در تشخیص هسته دارای ویژگی ۸۲/۵ ، حساسیت ۸۴ و دقت ۹۴/۲ و در تشخیص سیتوپلاسم دارای ویژگی ۷۸/۰۲، حساسیت ۷۹ و دقت ۹۲/۳ بود. تصاویر دارای وضوح ۷۶۸*۵۷۶ بود که توسط یک میکروسکوپ نوری معمولی با سه لنز چشمی اخذ شد و توسط یک کارت Pinnacle به منظور دیجیتال‌سازی به یک CCD متصل شده بود. با توجه به اثربخشی نتایج استخراج هسته و سیتوپلاسم در تشخیص نهایی، نتایج حاصل از این دو الگو به‌صورت جداگانه به‌دست آمده بود.

در پژوهش دیگری [9] سامانهٔ پیشنهادی با استفاده از ویژگی‌های تصاویر میکروسکوپی به بررسی تغییراتی مانند بافت، هندسه، رنگ و تجزیه و تحلیل آماری از تصاویر پرداخت. این تغییرات به‌عنوان یک ورودی طبقه‌بندی‌شده استفاده شد. روش ارائه‌شده نشان داد که روش مورفولوژیکی خودکار برای شناسایی حاد لنفوسیتی لوسمی (ALL) توسط تصاویر میکروسکوپ از نمونهٔ خون چگونه است. ابتدا ارزیابی در مورد شاخص‌های مورفولوژیکی از سلول‌ها انجام و درنهایت طبقه‌بندی برای تشخیص لوسمی انجام شد.

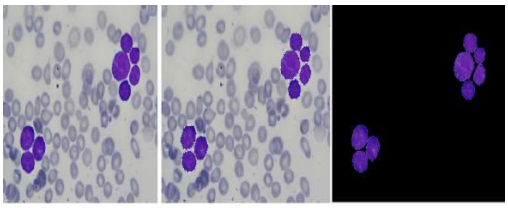
در مقاله‌ای [10] نمونه‌هایی از پنجاه بیمار سرطانی (لوسمی حاد لنفوییدی) در آزمایشگاه با لام‌های آماده تصاویر مورد نیاز از تمامی سلول‌ها تهیه شد. از این تصاویر به‌عنوان الگوهای لازم جهت آموزش شبکه‌های عصبی مصنوعی طراحی‌شده، استفاده شد. نرم‌افزاری متشکل از سه شبکهٔ عصبی مجزا و دو واحد تشخیص رنگ و اندازه‌گیری مساحت هسته طراحی شد. هر کدام از سه شبکهٔ عصبی به‌طور جداگانه و با استفاده از تصاویر تهیه‌شده مورد آموزش قرار داده شدند. این نرم‌افزار قادر به تشخیص گلبول‌های قرمز و سفید از یکدیگر بود.

در پژوهش دیگری [11] نشان داده شد که استخراج اطلاعات مربوطه از داده‌های Microarray کار بسیار پیچیده‌ای است که این ناشی از تعدد ویژگی‌هایی است که به‌طور طبیعی در داده‌های microarray وجود دارد. از این لحاظ، انتخاب ویژگی یک جنبه بسیار مهم از تجزیه و تحلیل است که در کارهای شناسایی ژن مربوطه و نیز برای بهبود دادن اطلاعات پیش‌بینی کمک می‌کند. در این مقاله، رافائل مارکوس و همکارانش در یک مطالعهٔ مقایسه‌ای بین انتخاب رو به جلو (SFS) و الگوریتم ژنتیک (GA) به‌عنوان چارچوب کلی برای تجزیه و تحلیل داده‌های Microarray و با هدف شناسایی

گروهی از ژن‌ها با قابلیت پیش‌بینی بالا و ارتباط بیولوژیکی انجام دادند. برای پیش‌بینی نتیجه سرطان از شش روش استاندارد و مبتنی بر یادگیری ماشین (تجزیه و تحلیل مشخص (LDA)، ماشین بردار پشتیبان (SVM)، ناو بیز (NB)، شبکه‌های عصبی C-MANTEC، نزدیکترین همسایه (KNN) و پرسپترون چندلایه (MLP) با استفاده از شش مجموعه داده عمومی استفاده شد. نتایج نشان داد پیش‌بینی نتیجهٔ سرطان با استفاده از GA در مقایسه با یک SFS بهتر جواب می‌دهد.

۳- روش پیشنهادی

در این مقاله ترکیبی از سه الگوریتم شامل الگوریتم FCM برای خوشه‌بندی تصاویر و الگوریتم ژنتیک برای بهبود نتایج و شبکهٔ عصبی برای دسته‌بندی و تشخیص استفاده شده است. شکل (۱) نشان‌دهنده روندنمای روش پیشنهادی است. در تشخیص بیماری‌ها با استفاده از پردازش تصویر سه مرحلهٔ اصلی شامل پیش پردازش، پردازش و تشخیص وجود دارد.



(شکل-۱): a) تصویر اولیه b) مرزبندی هسته c) جداسازی هسته
(Figure-1): a) Primary image b) The Image after core boundary c) Separation of the core

• پیش‌پردازش

در پردازش تصویر، استخراج ویژگی و حذف نوفه از تصاویر، بسیار مهم است. تصاویر نوفه‌دار همیشه باعث اختلال در روند می‌شوند؛ به همین دلیل حذف نوفه از تصویر یکی از مهم‌ترین مراحل در پردازش تصویر است چون باعث بهبود در روند تشخیص می‌شود. در مرحله پیش‌پردازش ابتدا تصویر ورودی از یک فیلتر میانه برای ازبین‌بردن نوفه‌های موجود عبور داده می‌شود تا یک تصویر نرم و واضح به‌دست آید.

• پردازش

تقسیم‌بندی هسته یکی از کلیدی‌ترین مراحل پردازش است. عملکرد این مرحله، استخراج ویژگی و تقسیم‌بندی را تحت تأثیر قرار می‌دهد. در اینجا هسته طی دو مرحله از تصویر جدا می‌شود. ابتدا مجموعه‌ای از هسته‌ها توسط الگوریتم FCM به‌دست آمده، سپس اشیای اضافی از خوشه‌ها حذف، و هسته‌ها جدا می‌شوند. الگوریتم FCM یکی از دقیق‌ترین الگوریتم‌ها دسته‌بندی موجود است. این خوشه‌بندی بر مبنای تفکیک رنگ‌ها انجام می‌شود؛ به‌طوری‌که هر پیکسل از یک شیء به چهار خوشه دسته‌بندی می‌شود. این خوشه‌ها شامل هسته، پس زمینه و سلول‌های دیگر (گلبول‌های قرمز و سیتوپلاسم) است. با توجه به اینکه مقاله حاضر بر مبنای الگوریتم FCM است در ابتدا به معرفی این الگوریتم پرداخته شده و در ادامه روش پیشنهادی تفصیل خواهد شد.

• خوشه‌بندی فازی c میانگین (FCM)

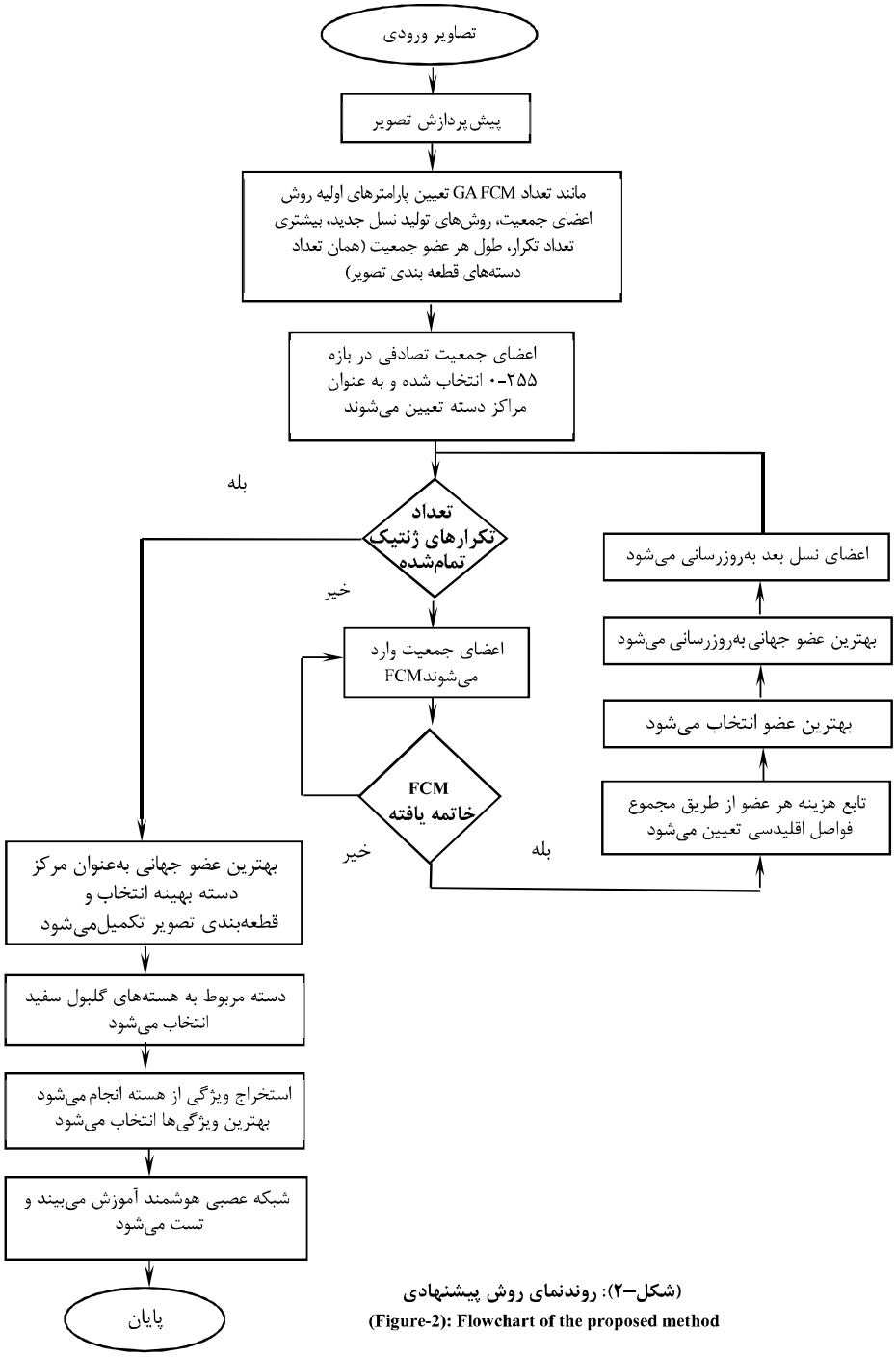
در خوشه‌بندی فازی هر نمونه ورودی می‌تواند متعلق به یک یا چند خوشه باشد و منحصر به یک خوشه نیست. در این الگوریتم نیز تعداد خوشه‌ها (c) از قبل مشخص شده است، شش دسته برای دسته‌بندی انتخاب شده که اعم از: L1، L2، L3، At، N و Re است. این دسته‌بندی‌ها با سه ویژگی ارزیابی شده‌اند که در بخش نتایج به آنها اشاره خواهد شد. تابع هدفی که برای این الگوریتم تعریف شده به‌صورت رابطه (۱) است که در آن m یک عدد حقیقی بزرگتر از یک است که در بیش‌تر موارد برای m عدد دو انتخاب می‌شود.

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} \| x_k - v_i \|^2 \qquad (۱)$$

اگر در این رابطه m را برابر یک قرار دهیم، تابع هدف خوشه‌بندی c میانگین (کلاسیک) غیر فازی به‌دست می‌آید. در این رابطه x_k نمونه k ام و v_i نماینده یا مرکز خوشه i ام و n تعداد نمونه‌ها و u_{ik} میزان تعلق نمونه i ام در خوشه k ام را نشان می‌دهد. علامت ||*|| میزان فاصله از مرکز خوشه است که از هر تابعی که بیان‌گر تشابه نمونه و مرکز خوشه بوده می‌توان استفاده کرد. از u_{ik} یک ماتریس U می‌توان تعریف کرد که دارای c سطر و n ستون است و مؤلفه‌های آن هر مقداری بین صفر تا یک را می‌توانند اختیار کنند. اگر تمامی مؤلفه‌های ماتریس U به‌صورت صفر و یا یک باشند، الگوریتم مشابه c میانگین کلاسیک خواهد بود.

مراحل کلی الگوریتم FCM به‌صورت زیر است:

۱. مقداردهی اولیه برای c، m و U0. خوشه‌های اولیه حدس زده شوند.
۲. مراکز خوشه‌ها محاسبه شوند (محاسبه viها).
۳. محاسبه ماتریس تعلق از روی خوشه‌های محاسبه‌شده در مرحلهٔ دو.
۴. اگر $\|Ul+1-Ul\| \le \varepsilon$ الگوریتم خاتمه می‌یابد و در غیر این‌صورت برو به مرحلهٔ دو.



(شکل-۲): روندنمای روش پیشنهادی

(Figure-2): Flowchart of the proposed method

شکل (۲) یک نمونه از تقسیم‌بندی هسته را نمایش می‌دهد. نتایج نشان می‌دهد که خوشه با کمینه رنگ قرمز، خوشه مربوط به هسته است. به همین دلیل مقدار متوسط قسمت R که از همان بخش‌بندی رنگ RGB است، برای هر خوشه محاسبه و خوشه با کمینه مقدار متوسط به‌عنوان مجموعه‌ای از هسته در نظر گرفته می‌شود. بعد از مشخص‌کردن هسته اگر اشیا یا لکه‌های غیر هسته وجود داشته باشند، حذف می‌شوند.

•استخراج و انتخاب ویژگی

این بخش شامل دو مرحله و به شرح زیر است: مرحله نخست استخراج "ویژگی‌های نسل" است که در آن مجموعه‌ای از ویژگی‌های هستهٔ خوشه‌بندی‌شده توسعه پیدا می‌کند. مرحله دوم "انتخاب ویژگی‌ها" است که در آن مجموعه‌ای مطلوب از ویژگی‌ها که منجر به به‌دست‌آمدن تشخیص بهینه است، انتخاب می‌شوند.

به منظور تمایز بین سلول‌ها، باید ویژگی‌های مناسب از سلول‌ها انتخاب شود. در این مقاله ویژگی‌های هندسی و آماری برای طبقه‌بندی سلول استفاده شده‌اند. به‌طور خلاصه ویژگی‌های هندسی دارای اطلاعات در مورد اندازه و شکل هسته‌ها و ویژگی‌های آماری دارای اطلاعات درباره هیستوگرام تصویر سطح خاکستری پیکسل در هسته‌ها است. کل ویژگی‌های استخراج‌شده برابر با پانزده ویژگی (F8-F15) است که با استفاده از روش مرسوم PCA[1] [12]، ۹ ویژگی مهم که در ذیل نام برده شدند جهت تشکیل بردار ویژگی ماتریس سطح خاکستری رخداد (GLCM)[2] به کار رفتند. (دلیل انتخاب این ویژگی‌ها در جدول (۳) قابل مشاهده است). ۹ ویژگی مربوط به ماتریس GCLM از تصویر Gayscale است که در زیر رابطه‌های ریاضی آنها آمده است.

$$\text{انرژی} = \sum_{i,j=0}^{n-1} \left(p_{i,j}\right)^2$$

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

$$\text{همبستگی} = \frac{\sum_i \sum_j (i,j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

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

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

$$\text{محیط} = 2 * pi * r$$

$$\text{قطر} = sqrt\left(4 * \frac{\text{سطح}}{pi}\right)$$

$$\text{عدد اولر} = e = \lim_{n\to\infty} [1 + \frac{1}{n}]^n$$

$$\text{استحکام} = \frac{\text{مساحت}}{\text{مساحت بدنهٔ محدب}}$$

شش ویژگی باقیمانده مربوط به آنتروپی بافت رنگی از فضاهای H^3، S^4 و V^5 در فضای HSV و R^6، G^7 و B^8 در فضای RGB استخراج شده است. به‌طورکلی ۹ ویژگی GCLM از فضای خاکستری و شش ویژگی از فضای رنگی استخراج شده است. رابطه آنتروپی به‌صورت زیر تعریف می‌شود:

$$\text{آنتروپی} = -\sum_{i,j} p(i,j) \times log_2 p(i,j)$$

• بهینه‌سازی خوشه‌بندی با استفاده از GA

نتیجهٔ به‌دست‌آمده تنها با استفاده از FCM در قطعه‌بندی تصویر، ممکن است، راضی‌کننده نباشد. بنابراین به‌منظور تشخیص سرطان خون، FCM و GA به‌ترتیب برای نایل‌شدن به بهترین نتیجه به‌کار می‌روند. نخست چرخه با خواندن تصویر پزشکی شروع می‌شود و سپس FCM از ویژگی قطعه‌بندی تصویر استفاده کرده و پیکسل‌ها را در گروه‌هایی سازماندهی می‌کند. هر گروه به‌عنوان یک مرکز خوشه استفاده می‌شود و این مرکز برای GA آماده شده تا راه حل بهینه در تشخیص تصویر انتخاب شود. این الگوریتم از اکنون تا انتهای مقاله GAFCM نام‌گذاری می‌شود.

در روند الگوریتم FCM که یک روند بهینه‌یابی کلاسیک محسوب می‌شود، هدف کمینه‌کردن پارامتر J در رابطه (۲) است.

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2} = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} \left\| x_k - v_i \right\|^2 \qquad (۲)$$

$$\mathrm{d_{ij}} = \left\| \mathrm{c_i - x_j} \right\| \; d_{ij} = \left\| c_i - x_j \right\| \qquad (۳)$$

[1] Principal Component Analysis
[2] Gray Level Co-Occurrence Matrix

[3] Hue
[4] Saturation
[5] Value
[6] Red
[7] Green
[8] Blue

همان‌طورکه در رابطه (۲) مشاهده می‌شود، این کمیت وابسته به d_{ij} ها است. طبق رابطه (۳) d_{ij} نیز وابسته مراکز خوشه‌هاست. حال اگر C_i ها به‌گونه‌ای انتخاب شوند که کمترین فاصله اقلیدسی با پیکسل‌های درون هر خوشه را داشته باشند، می‌توان گفت مقدار J را به‌صورت بهینه کاهش داده‌ایم. از این رو به جای استفاده از مرحله دو در الگوریتم FCM، در الگوریتم پیشنهادی وظیفه جایابی بهینه مقادیر C_i را به الگوریتم GA واگذار می‌کنیم. الگوریتم GA در هر بار تکرار مقادیر مختلفی از اعضای جمعیت را تولید کرده و این اعضا وارد روند FCM می‌شوند؛ سپس مقدار J از فرمول (۲) تعیین و عضوی که کمترین مقدار J را داشته باشد، به‌عنوان بهترین عضو انتخاب شده و تصویر با استفاده از این مراکز دسته و مقادیر تابع عضویت خوشه‌بندی می‌شود. به‌منظور دست‌یابی به مقدار بهینه در فرمول (۲) مشتق مرتبه نخست به‌صورت رابطه (۴) محاسبه می‌شود:

$$\frac{\partial J}{\partial c_j}=\frac{\partial}{\partial c_j}\sum_{i=1}^{n}(x_i-c_j)^2=-2\sum_{i=1}^{n}(x_i-c_j)=0 \qquad (۴)$$

که در نتیجه رابطه (۵) به‌دست می‌آید:

$$c_j=\frac{1}{n}\sum_{i=1}^{n}x_i,(\mathrm{j}=1,2,..k) \qquad (۵)$$

در این حالت پس از گذشت چند تکرار، الگوریتم به مقدار ثابتی هم‌گرا می شود و مراکز خو شه به‌گونه‌ای انتخاب می‌شوند که داده‌های هر خوشه بیشترین شباهت را به مرکز خوشه داشته باشند.

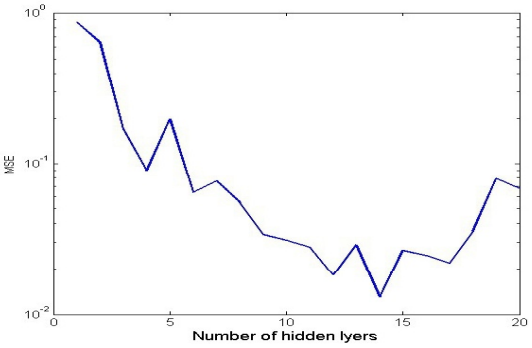
• طبقه‌بندی

آخرین مرحله در تشخیص سرطان خون این است که نتایج به‌دست آمده از تصویر در مراحل قبل را با استفاده از شبکه عصبی طبقه‌بندی کرد. این شبکه با استفاده از تصاویر سرطانی و غیر سرطانی گلبول‌های خون آموزش داده شد. به‌منظور یافتن تعداد مناسب لایه‌های پنهان در شبکه عصبی، یک شبکهٔ پرسپترون با الگوریتم آموزشی Gradient Descends ساخته شد و با افزایش تعداد لایه‌های پنهان از دو تا بیست نمودار رسم‌شده در شکل (۳) برای خطای MSE شبکه به‌دست آمد.

همان‌طورکه پیداست، تعداد لایه‌ها در چهارده کمترین MSE را دارد. از این رو شبکهٔ طراحی‌شده از نوع پرسپترون با چهارده لایهٔ پنهان در نظر گرفته شد.

محیط آزمایشگاهی

محیط آزمایشگاهی در این کار نرم‌افزار متلب ۲۰۱۱ است. آزمایش‌ها بر روی یک رایانه با پردازنده Core i5 و شش گیگابایت رم و ویندوز هفت انجام شده است. کل زمان پردازش‌شده برای انجام آزمایش برای هر عکس برابر با ۲۰ ثانیه است.



(شکل-۳): نمودار *MSE* برای ۲ تا ۲۰ لایه شبکه عصبی
(Figure-3): MSE Plot for 2 to 20 neural network layers

۴- نتایج

در شکل (۴) یک تصویر نمونه و تصویر حاصل از مرحله قطعه‌بندی نمایش داده شده است.

پارامترهای ارزیابی

جهت ارزیابی روش پیشنهادی پارامترهای حساسیت، ویژگی و دقت استفاده شده‌اند. بدین منظور ابتدا لازم است، مقادیر TP, TN, FP, FN استخراج شوند که عبارتند از:

TP: تعداد تصاویر سرطانی که به‌درستی سرطانی تشخیص داده می‌شوند، FP: تعداد تصاویر غیر سرطانی که به اشتباه سرطانی تشخیص داده می‌شوند، TN: تعداد تصاویر غیر سرطانی که به‌درستی سرطانی تشخیص داده نمی‌شوند، FN: تعداد تصاویر سرطانی که به اشتباه غیر سرطانی تشخیص داده می‌شوند.

پارامترهای ارزیابی طبق روابط ۶ تا ۸ قابل محاسبه هستند:

$$\text{حساسیت}=\frac{T_p}{T_p+F_n} \qquad (۶)$$

$$\text{ویژگی}=\frac{T_n}{T_n+F_p} \qquad (۷)$$

$$\text{دقت}=\frac{T_p+T_n}{T_p+T_n+F_p+F_n} \qquad (۸)$$

(جدول-۱): تعداد تصاویر به‌ازای هر پارامتر
(Table-1): The number of images in terms of different Parameters

شماره ویژگی	حساسیت	ویژگی	دقت
F1	40.1	81.47	73.02
F2	41.05	81.59	73.58
F3	41.09	81.48	73.68
F4	41.1	82	73.98
F5	41.3	82.03	74.22
F6	41.5	82.06	74.35
F7	41.6	82.2	74.65
F8	41.9	82.41	74.89
F9	42.8	82.96	75.2
F10	44.1	82.99	75.42
F11	45.89	83	75.88
F12	45.94	83.22	76.1
F13	46.7	83.14	76.2
F14	48.6	83.86	76.55
F15	49.85	84.01	76.87

تعداد تصاویر مورد استفاده در این مقاله برای تصاویر دارای تومور برابر با ۲۱ و برای تصاویر غیر تومور ۲۳ است.

(جدول-۲): نتایج قطعه‌بندی روش GAFCM
(Table -2):The Result of GAFCM Segmentation

TP	FN	FP	TN
۲۰	۱	۱	۲۲

جدول (۲) نتایج خوشه‌بندی را در قالب پارامترهای ارزیابی حساسیت، ویژگی و دقت به‌ازای پارامترهای آماری مختلف نشان می‌دهد. L1, L2 ,L3 به سه زیر شاخه مورفولوژیکی در تشخیص سرطان خون ALL اشاره دارد و حالت‌های active, normal , reactive نشان می‌دهد که چند درصد از موارد شناسایی‌شده مربوط به بیماران با سرطان فعال، بدون سرطان و حالت بازگشت سرطان است. حالت Mean میانگین کلی پارامترهای ارزیابی‌شده را نشان می‌دهد.

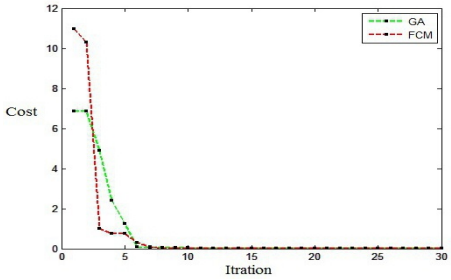
همان‌گونه‌که از جدول (۲) مشخص است، نتایج حاصل از آزمایش‌ها نشان می‌دهد که روش GAFCM نسبت به روش‌های دیگر از نظر سه مشخصه حساسیت، دقت و ویژگی کارامدتر است.

جدول(۳) نتایج حاصل از طبقه‌بندی با ویژگی‌های مختلف را به‌ازای سه مشخصه از جمله، حساسیت، ویژگی و دقت مورد بررسی قرار داده است. همان‌طورکه مشاهده می‌شود، مقدار ویژگی گزارش‌شده به‌ازای تمام ویژگی‌های استخراج‌شده نسبت به مقدار دقت و حساسیت بیشتر است. در این جدول F ها به ویژگی‌های استخراج‌شده از طریق ماتریس GLCM اشاره دارند. در این آزمایش درمی‌یابیم که مهم‌ترین ویژگی‌های مورد نیاز ما که توسط PCA مرتب‌سازی شده‌اند، ویژگی‌های F8 تا F15 است.

(جدول-۳): نتایج حاصل از طبقه‌بندی با ویژگی‌های مختلف
(Table-3): Results of classification for different Features

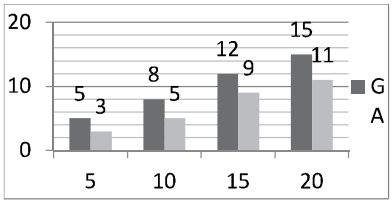
پارامترها	L1	L2	L3	Active	Normal	Reactive	Mean
حساسیت	91.0	85.1	96.9	97.0	68.5	72.45	85.15
ویژگی	98.4	96	98.7	99	97.23	99.74	98.17
دقت	96	91.2	98.7	98.44	96.88	98	96.53

نمودار هزینه الگوریتم GAFCM در هر تکرار سنجیده شده و در شکل (۴) نشان داده شده است. نمودار سبزرنگ در این شکل بیان‌گر الگوریتم GAFCM و نمودار قرمز نشان‌گر الگوریتم FCM است. به‌طورکامل مشخص است که سرعت هم‌گرایی FCM از الگوریتم ژنتیک بیشتر، اما میزان کاهش هزینه در الگوریتم GAFCM بیشتر است.



(شکل-۴): نمودار هزینه برای هر دو الگوریتم
(Figure-4): Plot of Cost for both algorithms

علاوه‌براین، سرعت اجرای هر دو الگوریتم، اندازه‌گیری و در شکل (۵) نمایش داده شده است.



(شکل-۵): نمودار سرعت اجرای برنامه به‌ازای تعداد تصاویر مختلف
(Figure-5): execution time of algorithms for different Images

مشخصه‌های دقت، ویژگی و حساسیت در روش GAFCM با سه روش اخیر در همین حوزه مقایسه شده است. نتایج این مقایسه در جدول (۴) مشاهده می‌شود.

(جدول-۴): مقایسه روش پیشنهادی با روش‌های پیشین
(Table-4): Comparison of the proposed and previous method

روش	حساسیت	ویژگی	دقت
GAFCM	٪۸۵/۱۵	٪۹۸/۱۷	٪۹۶/۵۳
(Moradi et al, 2015)	٪۸۴/۳	٪۹۷/۳	٪۹۵/۶
(soltanzadeh et al, 2010)	٪۸۴	٪۸۲/۵	٪۹۴/۲
(Halim et al, 2011)	٪۸۴/۵	٪۹۷/۸	٪۹۴/۳

همان‌طورکه در جدول (۴) مشخص است، روش ترکیبیGAFCM توانسته با دقت، حساسیت و ویژگی بالاتری به تشخیص سرطان خون از روی تصاویر سلول‌های خونی منجر شود.

۵- نتیجه‌گیری

انواع لوسمی بر اساس سرعت پیشرفت بیماری و بدترشدن وضعیت بیماری، گروه‌بندی می‌شود. لوسمی به دو صورت مزمن (بیماری به آرامی پیشرفت می‌کند) یا حاد (بیماری سریع پیشرفت کرده و بدتر می‌شود) است. در این مقاله با استفاده از ترکیب الگوریتم‌های ژنتیک و FCM، الگوریتم جدید GAFCM پیشنهاد شد که در بخش خوشه‌بندی تصویر نتایج قابل قبولی به همراه داشته و از شبکه عصبی برای دسته‌بندی سرطان‌ها بهره گرفته است. الگوریتم FCM یکی از الگوریتم‌های بهینه در خوشه‌بندی است؛ اما به‌دلیل ضعف آن در به دام افتادگی کمینه‌های محلی، در زمینه قطعه‌بندی هسته‌های سرطانی دچار مشکل می‌شود. ترکیب روش FCM و ژنتیک، موجب می‌شود که وظیفه بهینه‌یابی مراکز خوشه در الگوریتم FCM به روش قدرتمند ژنتیک محول شود. نتایج شبیه‌سازی روش پیشنهادی و مقایسه آن با دیگر روش‌های تشخیص سرطان خون نشان داد که روش پیشنهادی از درصد بالایی از پارامترهای حساسیت، ویژگی و دقت نسبت به روش‌های پیشین برخوردار است.

۶- مراجع 6- References

[1] Amin MM, Kermani S , Talebi A, Oqhli MG. Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K-Means Clustering and Support Vector Machine Classifier J Med Signals Sens. 2015 Jan-Mar;5(1):49-58.

[2] Dan L. Longo, Anthony S. Fauci, Dennis L. Kasper, Stephen L. Hauser, Larry Jameson, Joseph Loscalzo. Harrison's Principles of Internal Medicine, 18th Edition, Chapter 110. Malignancies of Lymphoid Cells. Clinical Features, Treatment, and Prognosis of Specific Lymphoid Malignancies. 2015.

[3] Collier, J.A.B Oxford Handbook of Clinical Specialties, Third Edition. Oxford. 1991; pp. 810. ISBN 0-19-262116-5.

[4] ACS : How Is Acute Lymphocytic Leukemia Classified?". 2016. [Available from:http:// www. Cancer. org/docroot/CRI/ content/ CR-I_2_4_3X_ How_ Is_ Acute_ Lymphocytic_ Leukemia_ Classified. asp? rnav= cri].

[5] Hoffbrand AV, Moss PAH and Pettit JE. "Essential Haematology", Blackwell, 5th ed., 2006.

[6] Messinger YH, Gaynon PS, Sposto R, van der Giessen J, Eckroth E, Malvar J, et al. Therapeutic Advances in Childhood Leukemia & Lymphoma (TACL) Consortium. "Bortezomib with chemo-therapy is highly active in advanced B-precursor acute lymphoblastic leukemia: Thera-peutic Advances in Childhood Leukemia & Lymphoma (TACL) Study". Blood. 2012; 120 (2): 285–90.

[7] Lambrou GI, Papadimitriou L, Chrousos GP, Vlahopoulos SA. "Glucocorticoid and pro-teasome inhibitor impact on the leukemic lymphoblast: multiple, diverse signals converg-ing on a few key downstream regulators". Mol Cell Endocrinol. 2012; 351 (2): 142–51.

[8] Halim NH, Mashor MY, Hassan R. Automatic blasts counting for acute leukemia based on blood samples. Int J Res Rev Comput Sci 2:971, 2011.

[9] Heydari H, Souratgar A.A, Rashidi I, Malekpoor N, Parvizi A. Diagnosis of leukemia [a particular type of acute lymphoblastic leukemia by using artificial neural networks. Iranian Student Conference on Electrical Engineering, Tarbiat Modarres University. 24- 26 September 1389.

[10] Luque-Baena RM, Urda D, Subirats JL, Franco L, Jerez JM. Application of genetic algorithms and constructive neural networks for the analysis of microarray cancer data. Theoretical Biology and Medical Modelling 2014, 11(Suppl 1):S7.

[11] Soltanzadeh R, Rabbani H. Talebi A. Classification of Three Types of Red Blood Cells in Peripheral Blood Smear in Proc. IEEE Int. Conf. on Signal Processing, pp. 707 – 710, China, 2010.

[12] Moradi P, Ahmadian S, Akhalghian F. An effective trust-based recommendation method using a novel graph clustering algorithm. Statistical Mechanics and its Applications. Volume 436, 15 October 2015, Pages 462–481

[13] ZohourParvaz F, Fatemizadeh E, Behnam H. Speed improvement in graph-cuts-based registra-tion for non-rigid image registration of brain magnetic resonance images. JSDP. 2017; 13 (4) :79-92

عباس کریمی دکترای سامانه‌های امنیت بحرانی، فوق دکترای پردازش تصاویر پزشکی، فوق دکترای شبکه‌های موبایل هوشمند (از دانشگاه UPM)، کارشناسی ارشد مهندسی کامپیوتر نرم‌افزار، کارشناسی مهندسی کامپیوتر سخت‌افزار، عضو هیأت علمی ومدیر دپارتمان مهندسی کامپیوتر دانشگاه آزاد اسلامی اراک، مدیرعامل شرکت پژوهشی دانش‌بنیان IDAP، IEEE Senior Member، عضو هیأت تحریریه و ادیتور چندین ژرنال بین‌المللی WOS، عضو انجمن کامپیوتر ایران، با زمینه‌های پژوهشی؛ سامانه‌های تحمل‌پذیر خطای بلادرنگ، پردازش تصاویر پزشکی، امنیت و مانیتورینگ شبکه‌های سیار، شبکه‌های اجتماعی، ارزیابی کارایی سامانه‌های رایانه‌ای است.
نشانی رایانامهٔ ایشان عبارت است از:
Akarimi@iau-arak.ac.ir

لیلاسادات حسینی کار شناس ار شد مهندسی رایانهٔ نرم‌افزار دانشگاه آزاد اسلامی اراک، عضو انجمن کامپیوتر ایران، مدیر بخش پژوهش‌ها و توسعه سامانه‌های هوشمند تشخیص سرطان شرکت بیوانفورماتیک IDAP و با زمینه‌های پژوهشی بیوانفورماتیک پزشکی، سامانه‌های مانیتورینگ پزشکی است.
نشانی رایانامهٔ ایشان عبارت است از:
Rahenorayaneh@yahoo.com