مروری بر بخش بندی تصاویر CT ریه بیماران کووید-19 با استفاده از الگوریتم FCM

راضیه استکی 1\*، سید هاشم طبسی 1 و امین علیدوست جهزدانی1

1 دانشکده ریاضی و علوم کامپیوتر، دانشگاه دامغان، دامغان، ایران.

raziyeesteki@gmail.com

چکیده: بمنظور پیشگیری از پیشرفت اثرات بیماری کووید-19 که در حال حاضر داروی موثری برای درمان آن وجود ندارد، یکی از مهمترین راهکار ها، تشخیص مبنلایان در اسرع وقت است. بیشتر اثر ویروس بر ریه بیماران است، که بررسی به موقع و دقیق تصاویر ریه اهمیت زیادی دارد. با توجه به پیدایش مکرر سویه های گوناگون بیماری و افزایش ناگهانی تعداد بیماران، بررسی تصاویر ریه بیماران توسط پزشکان نیاز به زمان زیادی دارد و احتمال خطای تشخیص را افزایش می‌دهد. باتوجه به پیشرفت هوش مصنوعی در حوزه های پزشکی و کاربرد الگوریتم های یادگیری ماشین در آن، بررسی تصاویر پزشکی به کمک آن ها نقش مهمی در کنترل و پیشگیری این همه گیری ایفا می‌کند. در مطالعه حاضر با استفاده از الگوریتم Fuzzy C-means، تصاویر CT ریه بیماران مورد بررسی قرار گرفته است.

**کلید واژه‌ها:** الگوریتم FCM، بخش بندی، تصاویر CT، ریه، کووید-19.

1- مقدمه

بیماری کووید-19 یک بیماری مسری است که توسط سندرم حاد تنفسی ویروس کرونا (2SARS-CoV-) ایجاد می شود. اولین مورد شناخته شده در دسامبر 2019 در ووهان چین شناسایی شد [1]. براساس گزارشات منتشر شده از سوی سازمان جهانی بهداشت، این بیماری تا پایان ماه نوامبر 2021(ماه آبان 1400 خورشیدی). بیش از 260 میلیون نفر مبتلا و افزون بر 5 میلیون نفر تلفات داشته است.

از بین افرادی که علائم بیماریشان آنقدر قابل توجه است که به عنوان بیمار طبقه بندی می شوند، اکثر آنها (81٪) علائم خفیف تا متوسط (مانند ذات الریه خفیف) دارند، در حالی که 14٪ علائم شدید را نشان می دهند(مانند تنگی نفس، کمبود میزان اکسیژن خون یا بیش از 50 درصد درگیری ریه) و 5٪ از علائم بحرانی (نارسایی تنفسی، شوک، و اختلال عملکرد چند اندام) رنج می برند [2].

امروزه تقسیم بندی تصاویر به یکی از مهمترین ابزارهای پزشکی تبدیل شده است. بررسی تصاویر سی تی ریه برای مدیریت زودهنگام این بیماری مهم است. تصاویر سی تی قفسه سینه عملکرد بسیار ارزشمندی در شناسایی درگیری آسیب شناسی های مؤثر بر ریه ها دارند. تقسیم بندی تصویر پزشکی ابزار مهمی برای کاربردهای بالینی فعلی است. در پردازش تصویر دیجیتال و بینایی کامپیوتری، بخش بندی تصویر فرآیند تقسیم‌بندی تصویر دیجیتال به بخش‌های متعدد است. هدف تقسیم‌بندی ساده‌سازی یا تغییر بازنمایی یک تصویر به چیزی است که معنادارتر و تحلیل آن آسان‌تر باشد [3]، [4]. تقسیم بندی تصویر معمولاً برای تعیین موقعیت اشیا و مرزها (خطوط، منحنی ها و غیره) در تصاویر استفاده می شود. به‌طور دقیق‌تر، تقسیم‌بندی تصویر، فرآیند تخصیص یک برچسب به هر پیکسل در یک تصویر است به طوری که پیکسل‌هایی با برچسب یکسان ویژگی‌های خاصی را به اشتراک می‌گذارند. تکنیک‌های مختلفی برای تقسیم‌بندی تصویر مانند آستانه، مبتنی بر لبه، مبتنی بر خوشه، مبتنی بر شبکه عصبی وجود دارد. از تکنیک های مختلف یکی از کارآمدترین روش ها روش خوشه بندی است.الگوریتم های متعددی برای بخش بندی تصاویر پزشکی از جمله الگوریتم های K-means ,Fuzzy C-Means ،EnFCM [5] و... ارائه شده اند که در این مطالعه به بررسی الگوریتم FCM پرداخته شده است.

هنگام انجام یک پروژه بینایی کامپیوتر، چیزی که همیشه به آن نیاز داریم داده است. در این مورد، داده ها تصاویر هستند. متأسفانه، چند مشکل مرتبط با داده های تصویر شامل پیچیدگی و عدم دقت کافی است. به همین دلیل قبل از ساخت یک مدل بینایی کامپیوتری، ضروری است که داده ها از قبل پردازش شوند تا به نتایج دلخواه برسیم. سپس کمیت های ویژگی مانند رنگ، روشنایی و مشخصات اطلاعات شکل تصاویر قابل مشاهده را انتخاب کنیم. در این مطالعه تصاویر با تغییر به اندازه استاندارد 512 × 512 و پس از رفع نویز آماده استفاده در الگوریتم بخش بندی تصاویر شدند. در ادامه با استفاده از تصاویر آماده شده در محیط برنامه نویسی پایتون، الگوریتم FCM پیاده سازی شده است.

2- پیشینه پژوهش

خوشه‌بندی یکی از تکنیک های یادگیری ماشین بدون نظارت است که جمعیت را به چند گروه یا خوشه تقسیم می‌کند به طوری که نقاط داده در یک گروه مشابه یکدیگر هستند و نقاط داده در گروه‌های مختلف متفاوت هستند. تاکنون روش های گوناگونی برای بخش بندی تصاویر ارائه و در این مطالعه دو الگوریتم k-means و FCM معرفی شده است.

2-1- الگوریتم K-means

الگوریتم K-means یک الگوریتم تکراری است که سعی می‌کند مجموعه داده را به خوشه‌های متمایز غیر همپوشانی از پیش تعریف‌شده تقسیم کند که در آن هر داده فقط به یک خوشه تعلق دارد [6]. این الگوریتم سعی می‌کند تا نقاط داده درون خوشه‌ای را تا حد ممکن شبیه کند و در عین حال خوشه‌ها را تا حد ممکن متفاوت نگه دارد. داده را به یک خوشه اختصاص می دهد به طوری که مجموع مجذور فاصله بین نقاط داده و مرکز خوشه حداقل باشد. هر چه تنوع کمتری در داخل خوشه ها داشته باشیم، نقاط داده در همان خوشه تشابه بیشتری دارند. هدف از خوشه‌بندی K-Means به حداقل رساندن واریانس کل درون خوشه‌ای یا تابع خطای مربعی [7] است.

|  |  |
| --- | --- |
| (1) |  |

که در آن k تعداد خوشه ها و n همه داده ها است.

روش کار الگوریتم k-means به شرح زیر است:

1. داده ها را در گروه های k دسته بندی می کند که k از قبل تعریف شده است.
2. Kنقطه را به طور تصادفی به عنوان مراکز خوشه انتخاب می شود.
3. طبق تابع فاصله اقلیدسی، اشیاء را به نزدیکترین مرکز خوشه خود اختصاص داده می شوند.
4. میانگین همه اشیاء در هر خوشه محاسبه می شود.
5. مراحل 3،2 و 4 تا رسیدن به نقاط مشابهی در هر خوشه در دورهای متوالی تکرار شود.

2-2- الگوریتم Fuzzy C-means

خوشه بندی فازی (خوشه بندی نرم یا k-means نرم ) نوعی خوشه بندی است که در آن هر داده می تواند به بیش از یک خوشه تعلق داشته باشد. درجات عضویت به هر یک از داده ها اختصاص داده می شود. این درجات عضویت نشان دهنده میزان تعلق داده به هر خوشه است. بنابراین، داده های لبه خوشه، با درجه های عضویت پایین تر، ممکن است به میزان کمتری نسبت به نقاط مرکز خوشه در خوشه قرار گیرند[8].

الگوریتم c-means فازی بسیار شبیه به الگوریتم k-means است:

* تعداد خوشه ها انتخاب شوند.
* ضرایبی را به طور تصادفی به هر داده برای قرار گرفتن در خوشه ها اختصاص داده شوند.
* تکرار الگوریتم تا همگرایی ادامه یابد.
* مرکز برای هر خوشه محاسبه شود.
* برای هر داده، ضریب تعلق به خوشه ها محاسبه شود.

هر داده x دارای مجموعه ای از ضرایب است که درجه تعلق به k امین خوشه wk *(x*) را نشان می دهد. در c-means فازی، مرکز خوشه میانگین تمام نقاط است بگونه ای که وزن آنها بر اساس درجه تعلق به آن خوشه است.

|  |  |
| --- | --- |
| (2) |  |

که m ابرپارامتری است که میزان فازی بودن خوشه را کنترل می کند. هر چه بالاتر باشد، در پایان خوشه فازی تر خواهد بود.

هدف در FCM به حداقل رساندن تابع هدف است:

|  |  |
| --- | --- |
| (3) |  |

که در آن

|  |  |
| --- | --- |
| (4) |  |

3- روش تحقیق

3-1- پیش پردازش

پیش پردازش تصویر اصطلاحی برای عملیات روی تصاویر در پایین ترین سطح انتزاع است. این عملیات محتوای اطلاعات تصویر را افزایش نمی دهد، اما اگر آنتروپی یک معیار اطلاعاتی باشد، آن را کاهش می دهد. هدف از پیش پردازش، بهبود داده‌های تصویر است که ویژگی‌های تصویر مربوط به پردازش و تحلیل بیشتر را افزایش می‌دهد چندین تکنیک برای پیش پردازش داده های تصویر مانند دگرگونی های هندسی، فیلتر کردن، تبدیل فوریه و بازسازی تصویر، تغییر اندازه تصویر، تبدیل تصاویر به مقیاس خاکستری و افزایش کنتراست استفاده می شود. افزایش کنتراست یک حوزه مهم در پردازش تصویر برای بینایی انسان و کامپیوتر است که به طور گسترده ای برای پردازش تصویر پزشکی و به عنوان مرحله پیش پردازش در تشخیص گفتار، سنتز بافت و بسیاری دیگر از برنامه های پردازش تصویر استفاده می شود. برخی از ابزارها و پلتفرم های مورد استفاده در پیش پردازش تصویر عبارتند ازPython ، Pytorch، OpenCV، Keras، Tensorflow وPillow .

|  |  |
| --- | --- |
|  |  |
|  |  |
| ب | الف |
| شکل 1- تصاویر قبل و بعد از پیش پردازش. (الف) قبل از پیش پردازش. (ب) پس از پیش پردازش | |

3-2- اجرای الگوریتم FCM

تعیین تعداد بهینه خوشه ها همواره یکی از چالش‌های اصلی پیاده سازی الگوریتم های خوشه بندی مبتنی بر تقسیم بندی است. در این الگوریتم ها پاسخ قطعی برای تعیین تعداد بهینه خوشه ها وجود ندارد. تعیین تعداد بهینه خوشه ها، از جهتی تابع پارامترهای تنظیم شده در الگوریتم است و از سوی دیگر تابع معیار به عنوان معیار شباهت بین داده ها در الگوریتم خوشه درنظر گرقته شده است. [9].

روش های تعیین تعداد بهینه خوشه ها به دو دسته کلی روش های مستقیم و روش های مبتنی بر آزمون های آماری تقسیم می‌شوند:

روش های مستقیم تعیین تعداد بهینه خوشه ها: این روش ها به دنبال بهینه سازی یک معیار به‌خصوص، مانند مجموع مربعات فواصل درون خوشه ای هستند. از جمله این متدها می‌توان به متد elbow و روش های مبتنی بر معیار silhouette اشاره کرد.

روش های مبتنی بر آزمون های آماری در تعیین تعداد بهینه خوشه ها: این متدها به دنبال تطبیق مشاهدات با فرض صفر-یک آزمون آماری هستند. از جمله این روش ها می‌توان به Gap Statistics اشاره کرد [10].

پارامتر فازی m، در الگوریتم FCM یک پارامتر کلیدی است که می تواند به طور قابل توجهی بر نتیجه خوشه بندی تأثیر بگذارد. شاخص اعتبار خوشه ای (CVI) نوعی تابع معیار برای اعتبارسنجی نتایج خوشه بندی است، در نتیجه تعداد خوشه بهینه یک مجموعه داده را تعیین می کند. از منظر اعتبارسنجی خوشه‌ای، روشی جدید برای انتخاب مقدار بهینه m در FCM پیشنهاد و از چهار CVI معروف، یعنی XB، VK، VT و SC برای خوشه‌بندی فازی استفاده شده است. در این روش مقدار بهینه m زمانی تعیین می شود که CVI ها به حداقل مقادیر خود برسند. نتایج تجربی روی چهار مجموعه داده مصنوعی و چهار مجموعه داده واقعی نشان داده‌اند که محدوده m، [2، 3.5] و فاصله بهینه [2.5، 3] است [11]. پارامتر i، نشان دهنده ی بیشترین دفعات تکرار الگوریتم برای رسیدن به همگرایی است و پارامتر e، معرف آستانه برای بررسی همگرایی که اپسیلون نامیده شده است.

|  |
| --- |
|  |
| شکل 2- تعیین پارامتر های استفاده شده در الگوریتم |

مطابق شکل2، تعداد خوشه ها برابر 4 و پارامتر فازی 2 ، بیشینه تکرار 100 و مقدار اپسیلون 0.05 در نظر گرفته شده است.

3-3- نتایج

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
| ب | الف |
| شکل 1- تصاویر قبل و بعد از اجرای الگوریتم FCM. (الف) قبل از پردازش. (ب) تصاویر خوشه بندی شده | |

بمنظور بررسی میزان دقت خوشه بندی از روابط (5)،(6) و (7) استفاده می شود.

|  |  |
| --- | --- |
| (5) |  |
| (6) |  |
| (7) |  |

که در آن TPمثبت واقعی، TN منفی واقعی، FP مثبت کاذب و FN منفی کاذب است[12].

4- بحث و نتیجه گیری

پردازش تصاویر دیجیتال از بخش های مهم هوش مصنوعی است و کاربرد های فراوانی در پزشکی دارد. در این مطالعه نتایج الگوریتم FCM در بخش بندی اثرات بیماری کووید-19 بر ریه نشان داده شده است. استفاده از این الگوریتم که به افزایش دقت و کاهش زمان تشخیص بیماری می انجامد، کمک های شایان توجهی به کادر درمان بیمارستان ها می‌کند. امید است با آشنایی این روش های تشخیصی کارآمد، گامی مهم در پیشبرد اهداف علمی پژوهشی برداشته شود.

**مراجع:**

1. Novel Coronavirus (2019-nCoV): situation report, 22﻿ (Report). World Health Organization. February 11, 2020
2. "Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)". U.S. Centers for Disease Control and Prevention (CDC). 6 April 2020. Archived from the original on 2 March 2020. Retrieved 19 April 2020.
3. Linda G. Shapiro and George C. Stockman (2001): “Computer Vision”, pp 279–325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3
4. Barghout, Lauren, and Lawrence W. Lee. "Perceptual information processing system." Paravue Inc. U.S. Patent Application 10/618,543, filed July 11, 2003.
5. J Fan, DY Yau, AK Elmagarmid, WG Aref, Automatic image segmentation by integrating color-edge extraction and seeded region growing. IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 10(10), 1454–1466,2001.
6. Vattani, A."k-means requires exponentially many iterations even in the plane" (PDF), Discrete and Computational Geometry, 2011, 45 (4), 596–616.
7. Anil Z Chitade, Colour based image segmentation using k-means clustering, International Journal of Engineering Science and Technology, 2010,Vol. 2(10), 5319-5325.
8. A. S. Bhide, Priyankapatil, and S. Dhande, “Brain Segmentation Using Fuzzy C Means Clustering To Detect Tumour Region.” International Journal Of Advanced Research In Computer Science And Electronics Engineering, Apr. 2012 Vol. 1, pp. 85-90.
9. Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2014. "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set." Journal of Statistical Software 61: 1–36.
10. Kaufman, Leonard, and Peter Rousseeuw. 1990. Finding Groups in Data: An Introduction to Cluster Analysis.
11. Zhou, K., Fu, C. & Yang, S. Fuzziness parameter selection in fuzzy c-means: The perspective of cluster validation. Sci. China Inf. Sci. 57, 1–8 (2014). https://doi.org/10.1007/s11432-014-5146-0.
12. A. Padma, and R. Sukanesh “A Wavelet Based Automatic Segmentation of Brain Tumor in CT Images Using Optimal Statistical Texture Features.” International Journal of Image Processing (IJIP), Vol. 5, pp. 552 - 563, Dec. 2011.

A review of lungs CT images segmentation of Covid-19 patients using FCM algorithm

Raziye Esteki1\*, Seyed Hashem Tabasi 1, Amin AlidoostJehezdani1  
  
1 Faculty of Mathematics and Computer Scienc, Damghan University, Damghan, Iran.

raziyeesteki@gmail.com

*Abstract*—In order to prevent the progression of the complications of Covid-19 disease, for which there is currently no effective drug to treat, one of the most important strategies is to identify the patients as soon as possible. Most of the effects of the virus are on the lungs of patients, so timely and accurate examination of lung images is very important. Due to the frequent occurrence of various strains of the disease and the sudden increase in the number of patients, examination of patients' lung images by physicians requires a lot of time and increases the likelihood of misdiagnosis. Due to the development of artificial intelligence in the field of medicine and the application of machine learning algorithms in it, the study of medical images with their help plays an important role in controlling and preventing this epidemic. In this paper, using Fuzzy C-means algorithm, CT images of patients' lungs were examined.

Keywords—covid-19, CT images, FCM algorithm, lungs, segmentation.