

Handwritten Persian Word Recognition Using Centroid Sequence Freeman Chain Code (CSFCC) Feature and Enhanced Classification

Sara Khosravi¹, Abdolah Chalechale^{2*}

1- Computer Dept., Eng. Faculty, Razi University, Kermanshah, Iran.

2*- Computer Dept., Eng. Faculty, Razi University, Kermanshah, Iran.

¹khosravi_un@yahoo.com, ^{2*}chalechale@razi.ac.ir

Corresponding author address: Abdolah Chalechale, Faculty of Computer Engineering, Razi University, Kermanshah, Iran, Post Code: 6714414971.

Abstract- In this paper, a new method is proposed for offline handwritten Persian words recognition. The proposed method introduces the Centroid Sequence Freeman Chain Code (CSFCC) as a new and powerful feature along with the use of morphological features and an optimized support vector machine (SVM) classifier. A combination of particle swarm optimization (PSO) and gravitational search algorithm (GSA), abbreviated to PSOGSA, has been employed to optimize the SVM classifier. In the proposed method, all the connected components of a word are detected and combined with each other. For this purpose, a pictorial dictionary of asymptomatic subwords has been made. In addition, a database has been created to include the positions of asymptomatic subwords in order to narrow down the search space and increase the speed and improve the recognition accuracy. Based on the position of a subword in a word, it is more likely to make the right decision and detect the subword, accurately. The proposed method was implemented on the Iranshahr Database, containing nearly 17000 images of handwritten names of 503 cities of Iran. The resultant recognition accuracy is 89% in the experiments, which shows the capability of the proposed method and improving the results, compared to the other well-known methods.

Keywords- Offline handwritten Persian word recognition, Centroid Sequence Freeman Chain Code (CSFCC), Morphological features, Support Vector Machines (SVM), PSOGSA optimization algorithm.

بازشناسی کلمات دست‌نویس فارسی براساس ویژگی کد زنجیره‌ای فریم‌دنباله

مرکز جرم و طبقه‌بند بهبود یافته

سارا خسروی^۱، عبدالله چاله‌چاله^{۲*}

۱- دانشکده مهندسی کامپیوتر، دانشگاه رازی، کرمانشاه، ایران.

*۲- دانشکده مهندسی کامپیوتر، دانشگاه رازی، کرمانشاه، ایران.

¹khosravi_un@yahoo.com, ^{2*}chalechale@razi.ac.ir

* نشانی نویسنده مسئول: عبدالله چاله‌چاله، کرمانشاه، خیابان دانشگاه، دانشگاه رازی، دانشکده مهندسی کامپیوتر، کد پستی: ۶۷۱۴۴۱۴۹۷۱

چکیده- در این مقاله روش جدیدی برای بازشناسی برون خط کلمات فارسی دست‌نویس ارائه شده است. روش پیشنهادی شامل معرفی ویژگی جدید و قدرتمند کد زنجیره‌ای فریم‌دنباله مرکز جرم (CSFCC)، در کنار استفاده از ویژگی‌های ریخت‌شناسی و همچنین استفاده از طبقه‌بند ماشین بردار پشتیبان (SVM) بهبود یافته می‌باشد. در روش پیشنهادی، طبقه‌بند ماشین بردار پشتیبان (SVM)، با استفاده از ترکیب دو الگوریتم بهینه‌سازی ازدحام ذرات (PSO) و الگوریتم جستجوی گراشی (GSA)، که به‌اختصار آن را PSOGSA می‌نامیم؛ بهبود یافته است. رویکرد به کار برده شده در این تحقیق، تشخیص تمامی اجزای متصل به هم در کلمه و سپس ترکیب آن‌ها با یکدیگر می‌باشد. به همین منظور، یک فرهنگ لغت تصویری از زیرکلمات بدون علامت تهیه شده است. علاوه بر این، به‌منظور کاهش فضای جستجو، افزایش سرعت و دقت بازشناسی، یک پایگاه داده از موقعیت زیرکلمات بدون علامت اصلی نیز ایجاد و مورد استفاده قرار گرفته است. این کار باعث شد؛ تا برای بازشناسی یک زیرکلمه، براساس موقعیت آن در کلمه اصلی تصمیم‌گیری و احتمال تشخیص درست افزایش یابد. روش پیشنهادی بر روی پایگاه داده شناخته شده‌ی ایران‌شهر که شامل حدود ۱۷۰۰۰ تصویر از اسامی دست‌نویس ۵۰۳ شهر ایران است؛ اجرا شده است. دقت بازشناسی به دست آمده در روش پیشنهادی، ۸۹٪ در بخش آزمایش می‌باشد؛ که نشان‌دهنده توانمندی روش پیشنهادی و بهبود نتایج در مقایسه با دیگر روش‌های موجود است.

واژه‌های کلیدی: بازشناسی برون خط کلمات فارسی دست‌نویس، کد زنجیره‌ای فریم‌دنباله مرکز جرم (CSFCC)، ویژگی‌های ریخت‌شناسی، ماشین بردار پشتیبان (SVM)، الگوریتم بهینه‌سازی PSOGSA.

۱- مقدمه

خواهد کرد. امروزه نرم‌افزارهای قدرتمندی در زمینه‌ی تشخیص و بازشناسی علائم، ارقام، حروف، متون و نوشتارهایی از این قبیل، در زبان‌های رایج دنیا مانند انگلیسی، فرانسوی، آلمانی و چینی طراحی و تولید شده است. اما برای بازشناسی کلمات فارسی/عربی، با توجه به اینکه تعداد زیادی از مردم جهان از این زبان استفاده می‌کنند، کارهای نسبتاً کم و پراکنده‌ای صورت گرفته، که نتایج به‌دست آمده در مقایسه با کارهای انجام‌شده در سایر زبان‌ها، چندان رضایت‌بخش نبوده است [۳]. از این‌رو، با توجه به سابقه کوتاه بازشناسی نوری کاراکتر در زبان فارسی/عربی، پیچیدگی و

تشخیص نوری کاراکتر (OCR)، یکی از قدیمی‌ترین شاخه‌های وابسته به بازشناسی الگو، هوش مصنوعی و بینایی ماشین محسوب می‌شود. به‌طور کلی، هدف و انتظار نهایی از یک سیستم OCR، رسیدن سیستم به سطح توانایی انسان در تشخیص و خواندن سریع متون می‌باشد [۱،۲]. با توجه به تحقیقات فراوانی که از دیرباز تاکنون در این زمینه صورت گرفته است؛ روز به روز کاربردهای بیشتری از سیستم‌های بازشناسی کاراکترها نمایان می‌شود و این به توسعه و تکامل بازشناسی الگو کمک بسیاری

مبتنی بر جداسازی کلمات به حروف سازنده‌شان هستند. استخراج ویژگی، عبارت است از تعیین ویژگی‌های دقیق از الگوهای هدف. انتخاب روش مناسب استخراج ویژگی، یکی از پیچیده‌ترین مراحل طراحی یک سیستم بازشناسی موفق است، زیرا نتایج حاصل از این مرحله به‌طور مستقیم بر روی مرحله بازشناسی تأثیر می‌گذارد [۹]. استخراج ویژگی در زمینه‌ی بازشناسی کاراکتر با روش‌های مختلفی مانند پروفایل‌گیری [۱۰]، ویژگی‌های ساختاری [۱۱]، تبدیل فوری [۱۲]، کد زنجیره‌ی محلی^۷، گرادیان، پروجکشن عمودی و افقی^۸ [۱۳]، ویژگی‌های آماری و هندسی [۱۴] و تبدیل موجک^۹ [۱۵] انجام و در بسیاری از موارد باعث بهبود نرخ بازشناسی شده است.

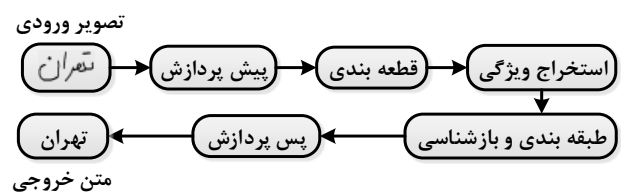
از طرفی، امروزه به دلیل افزایش اندازه پایگاه داده‌ها در کاربردهای مختلف، استفاده از مرحله انتخاب ویژگی در بازشناسی الگو در حال افزایش است. هدف از انتخاب ویژگی را می‌توان کاهش اندازه پایگاه داده با حفظ ویژگی‌های اصلی، حذف ویژگی‌های نامناسب، بهبود عملکرد یادگیری، کاهش زمان بازشناسی و افزایش دقت دانست [۹، ۱۶]. بیشتر تحقیقات صورت گرفته در این زمینه نیز در فرآیند انتخاب ویژگی، از الگوریتم‌های مبتنی بر جستجوی تصادفی مانند الگوریتم ژنتیک (GA) [۹] استفاده کرده‌اند. ویژگی‌های استخراج شده نیز شامل ویژگی‌های لحظه‌ای، ویژگی‌های مبتنی بر فاصله، ویژگی‌های هندسی و ویژگی‌های محلی می‌باشد. عزمی و همکاران در [۱۷]، با استفاده از روش ترکیبی الگوریتم‌های GA و شبیه‌سازی تبرید^{۱۰} (SA) به بازشناسی کاراکترهای فارسی دست‌نویس پرداختند. در مرحله اول نرخ بازشناسی با تعداد ۸۱ ویژگی، ۷۷ درصد و در مرحله دوم، بعد از اعمال الگوریتم ژنتیک که منجر به حذف برخی از ویژگی‌ها گردید؛ نرخ بازشناسی با تعداد ۵۵ ویژگی، تا ۸۰ درصد افزایش یافت. در مرحله سوم، با استفاده از الگوریتم ترکیبی GA و SA در مرحله انتخاب ویژگی، نرخ بازشناسی به ۸۲ درصد افزایش یافت.

طبقه‌بندی، مرحله تصمیم‌گیری اصلی یک سیستم OCR است. در این مرحله مرزهای صحیح بین بردارهای ویژگی طوری ایجاد می‌شوند که نمونه‌های هر الگو توسط مرزهای مشخص از دیگر نمونه‌ها جدا شوند [۱۳]. تکنیک‌های طبقه‌بندی بازشناسی کاراکتر را می‌توان به دو دسته‌ی تکنیک‌های کلاسیک و تکنیک‌های محاسباتی تقسیم‌بندی کرد. دقت عملکرد تکنیک‌های کلاسیک که شامل روش‌های ساختاری و روش‌های آماری و همچنین تطبیق الگو^{۱۱} است، به حجم داده‌ی مورد استفاده در تعریف پارامترهای مدل ساختاری بستگی دارد و با محدودیت‌های جدید سازگاری چندانی نداشته است، از این‌رو، اخیراً در این زمینه

دشواری بازشناسی این زبان به علت ویژگی‌های نگارشی منحصر به فرد نسبت به زبان‌های دیگر خصوصاً در متون دست‌نویس، نیاز به تحقیقات بیشتر در این حوزه احساس می‌شود. توجه به این نکته ضروری است که مدل زبان و شکل کاراکترهای آن نقش مهمی در میزان موفقیت بازشناسی دارد.

پیاده‌سازی یک سیستم بازشناسی کلمات دست‌نویس از دو جنبه اهمیت دارد. نخست اینکه هنوز متون بسیاری وجود دارند که نسخه الکترونیکی آن‌ها موجود نیست و ایجاد نسخه الکترونیکی آن‌ها با استفاده از یک سیستم بازشناسی قدرتمند می‌تواند نگهداری و دسترسی به این منابع را آسان نماید و از طرفی دیگر به علت سهولت نگارش با دست، هنوز افراد زیادی تمایل دارند که به جای تایپ متون، آن‌ها را دست‌نویس کنند. فراگیر بودن و وجود حجم عظیمی از اطلاعات کنونی روی کاغذ، محققان را بر آن داشته است که به دنبال روش‌هایی برای خواندن اتوماتیک این اطلاعات از روی کاغذ و تبدیل آن به شکل الکترونیکی باشند. برای تحقق این هدف سیستم بازشناسی کارآمدی لازم است تا متون دست‌نویس را به متون الکترونیکی تبدیل نماید.

همان‌گونه که در شکل ۱ نشان داده شده است، به‌طور کلی یک سیستم OCR شامل مراحل پیش‌پردازش، قطعه‌بندی^۱، استخراج ویژگی^۲، انتخاب ویژگی^۳، طبقه‌بندی^۴، بازشناسی^۵ و پس‌پردازش^۶ می‌باشد [۴]. لازم به ذکر است که با توجه به الگوریتم‌های مختلف بازشناسی و هدف مورد انتظار از آن‌ها، ممکن است در برخی از روش‌ها، بعضی از مراحل باهم ادغام و یا حذف شوند.



شکل ۱: مراحل اصلی یک سیستم OCR

طبقه‌بندی مرحله‌ی بسیار مهمی در سیستم‌های بازشناسی متون خصوصاً متون پیوسته می‌باشد. در نتیجه ضرورت انجام صحیح این مرحله و انتخاب مناسب روش طبقه‌بندی، در بالا بردن دقت سیستم بازشناسی کلمات فارسی از اهمیت بالایی برخوردار است. طبقه‌بندی متون را می‌توان به دو صورت کل‌نگر و تجزیه‌گرا انجام داد. در روش‌های کل‌نگر، بازشناسی براساس ویژگی‌ها و شکل عمومی کل کلمه، مستقل از حروف سازنده آن و بدون انجام عمل جداسازی، صورت می‌گیرد [۳، ۵]. در روش‌های تجزیه‌گرا، یک کلمه به کاراکترهای سازنده خود مانند زیرکلمات، حروف و یا شبه حروف، تجزیه شده و سپس مورد شناسایی قرار می‌گیرد. اکثر کارهای انجام شده در زمینه‌ی بازشناسی متون فارسی مانند [۸-۶]

مانند بازشناسی گفتار و یا متون کاربرد دارد. در [۲۶]، از مدل مخفی مارکوف، برای بازشناسی برخط حروف فارسی استفاده شده است. همچنین می‌توان از الگوریتم خوشه‌بندی فازی C -Means^{۲۱} (FCM)، برای یافتن مقدار بهینه عمومی مراکز خوشه‌ها و اختصاص درجه عضویت بردارهای ورودی به هر خوشه استفاده کرد. در [۲۳]، با استفاده از این روش، همراه با استراتژی تکاملی^{۲۲} (ES)، به بازشناسی کاراکترهای دست‌نویس فارسی پرداخته شده است. یکی دیگر از ساده‌ترین الگوریتم‌های یادگیری ماشین تحت نظارت، الگوریتم K -نزدیک‌ترین همسایه (K-NN) می‌باشد. این الگوریتم روشی غیر پارامتری است که نمونه‌ها را با توجه به رأی اکثریت طبقه‌بندی می‌کند. این نمونه‌ها متعلق به کلاس‌هایی در میان K -نزدیک‌ترین همسایه خود هستند. در [۲۴]، استفاده از ترکیب دو الگوریتم طبقه‌بندی KNN و شبکه عصبی برای بازشناسی کلمات دست‌نویس عربی برون خط و بدون محدودیت مورد بررسی قرار گرفته است. با توجه به نتایج به‌دست آمده، هر دو طبقه‌بند مورد نظر عملکرد نسبتاً بالایی در بازشناسی کلمات دست‌نویس عربی داشته‌اند.

در این مقاله به ارائه‌ی روشی جدید برای بازشناسی برون خط کلمات فارسی دست‌نویس، با استفاده از ویژگی معرفتی شده‌ی جدید و قدرتمند کد زنجیره‌ای فریمن دنباله مرکز جرم^{۲۳} (CSFCC) و همچنین استفاده از طبقه‌بند ماشین بردار پشتیبان (SVM) بهبودیافته با الگوریتم تکاملی PSOGSA پرداخته شده است. در روش پیشنهادی، بازشناسی بر مبنای زیرکلمات بدون علامت انجام شده و در مرحله استخراج ویژگی ابتدا با استفاده از بسته تبدیل موجک و سپس فیلتر کردن زیرباندهای حاوی اطلاعات غیرضروری داده‌های ورودی، تصاویر برای مرحله استخراج ویژگی آماده می‌شود. مرحله استخراج ویژگی شامل معرفتی ویژگی جدید کد زنجیره‌ای فریمن دنباله مرکز جرم (CSFCC) است که توصیف بسیار دقیقی از زیرکلمه دست‌نویس را فراهم می‌آورد. در کنار این ویژگی از ویژگی‌های ساختاری زیرکلمه نیز استفاده می‌شود. برای کاهش بار محاسباتی و در نتیجه افزایش سرعت و دقت بازشناسی، یک پایگاه داده از موقعیت زیرکلمات بدون علامت اصلی ایجاد کرده و بازشناسی زیرکلمات بر مبنای موقعیت آن‌ها در کلمه اصلی انجام می‌شود. در بخش طبقه‌بندی، پارامتر ضریب جریمه (C) و پارامتر کرنل (γ) الگوریتم SVM با استفاده از الگوریتم تکاملی PSOGSA بهینه می‌شود. در زیر نوآوری‌های مقاله به صورت موردی بیان گردیده است:

- معرفتی ویژگی جدید و قدرتمند کد زنجیره‌ای فریمن دنباله مرکز جرم (CSFCC)

تحقیقات گسترده‌ای صورت نگرفته است [۹]. از طرفی، در مطالعات صورت گرفته در دو دهه‌ی اخیر، استفاده از تکنیک‌های محاسباتی متنوعی مانند: ماشین بردار پشتیبان^{۱۲} (SVM) [۲۰-۱۸]، شبکه‌های عصبی^{۱۳} (NNs) [۹،۲۱]، شبکه‌های عصبی پیچشی^{۱۴} (CNN) [۲۲]، شبکه عصبی بازگشتی^{۱۵} (RNN) [۶]، منطق فازی [۲۳]، الگوریتم K -نزدیک‌ترین همسایه^{۱۶} (KNNs) [۲۴]، الگوریتم K میانگین^{۱۷} [۲۵] و مدل مخفی مارکف^{۱۸} (HMM) [۲۶] باعث افزایش سرعت بازشناسی، همگرایی سریع، یادگیری خودکار و انعطاف‌پذیری بالا در مواجهه با محدودیت‌های جدید، نسبت به تکنیک‌های کلاسیک شده و تحقیقات بسیاری در این حوزه انجام شده است [۹]. لازم به ذکر است که اکثر این الگوریتم‌ها مانند انواع شبکه‌های عصبی و همچنین الگوریتم SVM، جزء روش‌های یادگیری نظارت‌شده محسوب می‌شوند و برای آموزش به تعداد زیادی نمونه نیاز دارند.

شبکه عصبی پیچشی (CNN)، ترکیبی از یک استخراج‌گر ویژگی خودکار و یک طبقه‌بند قابل یادگیری است که از آن برای آموزش داده‌های پیچیده با ابعاد بالا استفاده می‌شود. در [۲۲]، با استفاده از ادغام دو طبقه‌بند شبکه عصبی پیچشی (CNN) و ماشین بردار پشتیبان (SVM) و با بهره‌گیری از تکنیک حذف تصادفی^{۱۹}، به بازشناسی برون خط کاراکترهای دست‌نویس عربی پرداخته شده است. در سیستم پیشنهادی مقاله [۲۲]، طبقه‌بند قابل یادگیری CNN، توسط طبقه‌بند SVM تصحیح می‌شود و از شبکه عصبی پیچشی برای استخراج اطلاعات ویژگی‌ها و از SVM به عنوان یک تشخیص‌دهنده (موتور بازشناسی)، استفاده شده است. استفاده از ترکیب دو یا چند طبقه‌بند با توجه به نوع مسئله، برای بهبود نتایج طبقه‌بندی و نرخ بازشناسی در یک مجموعه آموزشی یکسان و سپس ترکیب کردن نتایج آن‌ها برای عملکرد بهتر سیستم بازشناسی، هدف مورد نظر چندین پژوهش‌مانند [۱۳،۲۷،۲۸] بوده است. استفاده از طبقه‌بندهای گروهی که شامل چندین طبقه‌بند منحصر به فرد هستند، به بهبود نتیجه نهایی کمک می‌کند. به عبارتی دیگر، ایده‌ی اصلی یک طبقه‌بند گروهی، آموزش گروهی طبقه‌بندهای موازی و یا آبشاری، به جای یک طبقه‌بند است و از نتایج ترکیبی آن‌ها استفاده می‌کند. سیستم بازشناسی پیشنهاد شده در [۲۸] از چندین طبقه‌بند برای بازشناسی کلمات دست‌نویس عربی استفاده کرده است. طبقه‌های مورد استفاده در آن شامل پرسپترون چندلایه (MLP)، ماشین بردار پشتیبان (SVM) و طبقه‌بند یادگیری ماشین بی‌نهایت^{۲۰} (ELM) هستند.

مدل مخفی مارکوف (HMM) در مسائلی که بُعد زمانی دارند،

که در آن $rand$ عدد مستقل تصادفی تولید شده بین صفر و یک است. c_1 و c_2 ضرایب شتاب هستند و نحوه حرکت ذرات را کنترل می‌نمایند. ضریب وزنی (w) رفتار همگرایی الگوریتم را کنترل می‌نماید و باعث رسیدن به مقدار بهینه می‌گردد. در رابطه‌ی (۱)، عبارت wv_i^t قابلیت جستجو و اکتشاف را برای الگوریتم فراهم می‌آورد. همچنین $c_1 \times rand \times (pbest_i - x_i^t)$ و $c_2 \times rand \times (gbest - x_i^t)$ به ترتیب جزء شناختی و اجتماعی ذرات نامیده می‌شوند. در الگوریتم PSO ذرات به صورت تصادفی متناسب با فضای مسئله پخش می‌شوند. سپس بردار سرعت و موقعیت ذرات با استفاده از روابط (۱) و (۲) به‌روزرسانی شده و این فرایند تا زمان رسیدن به شرط پایان ادامه می‌یابد [۳۰].

۲-۲- الگوریتم جستجوی گرانشی (GSA)

الگوریتم GSA، که در سال ۲۰۰۹ توسط محققین ایرانی پیشنهاد شده است، یک روش بهینه‌سازی براساس قانون تعامل بین جاذبه و جرم می‌باشد. در GSA، عوامل به‌عنوان اشیا و عملکردشان توسط جرم آن‌ها بیان می‌شود. تمامی اشیا به‌وسیله‌ی نیروی گرانشی خود یکدیگر را جذب می‌کنند و این نیرو باعث ایجاد حرکت عمومی تمامی اشیا به سمت اشیایی با جرم سنگین‌تر می‌گردد [۳۱]. این الگوریتم، سیستم را به‌صورت مجموعه‌ای از N جرم در یک فضای n بُعدی در نظر می‌گیرد. موقعیت جرم، نقطه‌ای از فضا است که جوابی از مسئله می‌باشد. در این سیستم در زمان t نیروی وارد بر جرم i ام از سوی جرم j ام در جهت بُعد d ، به‌صورت رابطه (۳) تعریف می‌شود [۳۰]:

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \epsilon} \times (x_j^d(t) - x_i^d(t)) \quad (3)$$

که در آن $M_{aj}(t)$ جرم گرانشی فعال مربوط به جرم j ام و $M_{pi}(t)$ جرم گرانشی غیرفعال مربوط به جرم i ام است. $G(t)$ ثابت گرانشی در زمان t ام، ϵ مقدار ثابت بسیار کوچک برای جلوگیری از صفر شدن مخرج کسر و $R_{ij}(t)$ فاصله اقلیدسی بین دو جرم i ام و j ام است. مقدار $G(t)$ نیز مطابق با رابطه (۴) محاسبه می‌شود:

$$G(t) = G_0(t) \times \exp(-\alpha \times iter / maxiter) \quad (4)$$

که $iter$ و $maxiter$ به ترتیب تکرار فعلی و بیشترین تکرار الگوریتم می‌باشد. همچنین $G_0(t)$ ثابت گرانشی اولیه و α ضریب تضعیف است. کل نیرویی که بر جرم i ام در جهت بُعد d ام و در زمان t ام، وارد می‌شود؛ برابر با رابطه (۵) است:

$$F_i^d(t) = \sum_{j=1, j \neq i}^N rand_j F_{ij}^d(t) \quad (5)$$

- تشخیص کلمه‌ی ورودی براساس زیرکلمات بدون علامت
 - افزایش دقت بازشناسی از طریق بهینه کردن طبقه‌بند SVM با استفاده از الگوریتم PSO
 - افزایش سرعت و دقت بازشناسی و کاهش حجم شبکه با استفاده از چندین طبقه‌بند SVM با تنها یک خروجی
 - ایجاد پایگاه داده‌ی موقعیت زیرکلمات و بازشناسی زیرکلمه براساس موقعیت آن در کلمه اصلی
- بخش دوم مقاله به تشریح دو الگوریتم قدرتمند PSO و GSA و سپس ترکیب آن‌ها با یکدیگر می‌پردازد. بخش سوم به الگوریتم ماشین بردار پشتیبان و جزئیات آن مربوط می‌شود. در ادامه مقاله و در بخش چهارم، اجزای سیستم پیشنهادی معرفی خواهد شد. این بخش شامل توضیح در مورد پیش‌پردازش، چگونگی تهیه فرهنگ لغت تصویری زیرکلمات بدون علامت، قطعه‌بندی کلمات، ایجاد پایگاه داده موقعیت زیرکلمات، استخراج ویژگی‌ها با استفاده از روش پیشنهادی و درنهایت آموزش سیستم بازشناسی با استفاده از طبقه‌بند ماشین بردار پشتیبان بهبود یافته خواهد بود. بخش پنجم شامل معرفی پایگاه داده مورد استفاده در این تحقیق می‌باشد و در بخش ششم ارزیابی نتایج به‌دست آمده از روش پیشنهادی مورد بحث و بررسی قرار گرفته است. جمع‌بندی و نتیجه‌گیری کلی از مقاله نیز در بخش هفتم بیان شده است.

۲- الگوریتم‌های بهینه‌سازی فراابتکاری

هدف این الگوریتم‌ها به‌دست آوردن کمینه یک تابع با استفاده از جستجوی تصادفی در فضای پاسخ‌ها می‌باشد. در این قسمت به تشریح دو الگوریتم قدرتمند PSO و GSA و سپس ترکیب آن‌ها با یکدیگر پرداخته می‌شود.

۲-۱- الگوریتم بهینه‌سازی ازدحام ذرات (PSO)

الگوریتم PSO یک روش بهینه‌سازی قوی براساس هوش ازدحامی است که رفتار اجتماعی ذرات را شبیه‌سازی می‌کند. در این الگوریتم هر عضو به‌عنوان یک ذره بیان می‌شود و هر ذره به‌صورت بالقوه یک پاسخ مسئله می‌باشد [۲۹]. به عبارتی دیگر، هر ذره در تکرار t ام با بردار موقعیت x_i^t و بردار سرعت v_i^t یک پاسخ مسئله است. بهترین پاسخ به‌دست آمده توسط ذره i ام در تکرار t ام به‌عنوان $pbest_i$ و بهترین پاسخ در میان تمام ذرات، $gbest$ نامیده می‌شود. به‌روزرسانی سرعت و موقعیت در هر تکرار به ترتیب در رابطه‌های (۱) و (۲) آمده است [۳۰]:

$$v_i^{t+1} = wv_i^t + c_1 \times rand \times (pbest_i - x_i^t) + \dots \quad (1)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (2)$$

$$V_i(t+1) = w \times V_i(t) + c_1' \times rand \times ac_i(t) + \dots \quad (13)$$

$$c_2' \times rand \times (gbest - X_i(t)) \quad (14)$$

که در آن $V_i(t)$ ، سرعت ذره i ام در تکرار t ام، w ضریب وزنی، c_1' و c_2' ضرایب شتاب، $rand$ عدد مستقل تصادفی بین صفر و یک، $ac_i(t)$ شتاب ذره i ام در تکرار t ام، $X_i(t)$ بردار موقعیت جرم i ام در تکرار t ام و $gbest$ بهترین پاسخ در میان تمام جرم‌ها می‌باشد.

۳- الگوریتم ماشین بردار پشتیبان (SVM)

الگوریتم ماشین بردار پشتیبان، از یک تئوری یادگیری آماری گرفته شده است. در این تئوری، هدف جداسازی دو کلاس توسط یک صفحه جداساز بهینه است؛ به گونه‌ای که صفحه‌ی بهینه باید از داده‌های هر دو کلاس تا حد امکان فاصله داشته باشد. بردارهایی (نقاطی) که پهنا‌ی حاشیه^{۲۶} را محدود می‌کنند، بردارهای پشتیبان نامیده می‌شوند. در الگوریتم SVM، صفحات جداساز به گونه‌ای یافت می‌شوند؛ که حاشیه‌ی بین بردارهای پشتیبان حداکثر گردد. منظور از حاشیه، فاصله‌ی بین صفحه و نزدیک‌ترین نقطه‌ی داده‌ی هر کلاس با صفحه می‌باشد، [۱۸، ۲۲، ۳۲]. در صورتی که فضای ویژگی به صورت خطی باشد، فرض می‌شود که داده‌های آموزشی محدودیت‌های بیان شده در رابطه (۱۵) را برآورده کند [۳۲]:

$$\begin{cases} x_i^T w + b \gg +1 & \text{for } y_i = +1 \\ x_i^T w + b \ll -1 & \text{for } y_i = -1 \end{cases} \quad (15)$$

که در آن $y_i \in \{-1, +1\}$ و بیانگر برچسب الگوی آموزشی x_i است. صفحه‌ای که به صورت بهینه داده‌ها را از هم جدا می‌کند، به صورت رابطه (۱۶) تعریف می‌شود:

$$f(x) = \text{sign}(x^T w + b) \quad (16)$$

که x الگوی ورودی، w بردار وزن و b مقدار ثابت بایاس می‌باشد. مقدار وزن‌ها و بایاس توسط بیشینه کردن حاشیه $1/\|w\|$ برای N الگوی آموزشی، با حل مسئله بهینه‌سازی رابطه (۱۷) محاسبه می‌شود:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad y_i(x_i^T w + b) \gg 1, \quad i = 1, \dots, N \quad (17)$$

این مسئله قابل تبدیل به مسئله، بیشینه کردن دوگان لاگرانژ^{۲۷} مطابق رابطه (۱۸) است:

$$L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i^T x_j) \quad (18)$$

$$\alpha_i \geq 0, \quad i = 1, \dots, N, \quad \sum_{i=1}^N \alpha_i y_i = 0$$

که در آن α_i ضرایب لاگرانژ بوده و یک مسئله‌ی درجه دو استاندارد است و همواره دارای یک بیشینه عمومی α_i می‌باشد. مقدار w مطابق رابطه (۱۹) قابل محاسبه است:

که در آن $rand_j$ عددی تصادفی در بازه‌ی $[0,1]$ است. مطابق رابطه (۶)، هر جرم در جهت بُعد d ، شتابی می‌گیرد؛ که متناسب با نیروی وارده بر جرم i ام در آن جهت، بخش بر جرم i ام است.

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \quad (6)$$

سرعت بُعد d ام جرم i ام طبق رابطه (۷) محاسبه می‌شود:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \quad (7)$$

هنگامی که شتاب و سرعت هر جرم محاسبه شود، موقعیت جدید آن جرم می‌تواند توسط رابطه (۸) به روز شود:

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (8)$$

موقعیت جدید به معنای جرم‌های جدید است. جرم عامل‌ها بر مبنای تابع هدف تنظیم می‌شود، به گونه‌ای که به عامل‌هایی با شایستگی بیشتر، جرم بیشتری نسبت داده می‌شود. بنابراین جرم‌های گرانشی از طریق رابطه‌های (۹) و (۱۰) به روز می‌شوند:

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \quad (9)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)} \quad (10)$$

که $fit_i(t)$ بیانگر میزان برازندگی جرم i ام در زمان t ام و نیز مقدار $worst(t)$ و $best(t)$ به ترتیب بیانگر بدترین و بهترین مقدار برازندگی در کل جمعیت در زمان t ام هستند؛ و برای یک مسئله کمینه‌سازی به صورت روابط (۱۱) و (۱۲) تعریف می‌شوند:

$$best(t) = \min fit_j(t), \quad j \in \{1, \dots, N\} \quad (11)$$

$$worst(t) = \max fit_j(t), \quad j \in \{1, \dots, N\} \quad (12)$$

۳-۲ الگوریتم ترکیبی PSO-GSA

ایده اصلی الگوریتم ترکیبی PSO-GSA، ترکیب توانایی تفکر اجتماعی (g_{best}) در الگوریتم PSO و قابلیت جستجوی محلی الگوریتم GSA می‌باشد [۳۰]. توانایی الگوریتم GSA در جستجوی بهینه‌ی محلی بسیار خوب می‌باشد؛ اما سرعت جستجو در تکرارهای آخر کند است. از طرفی الگوریتم PSO سرعت مناسبی در یافتن بهینه‌ی عمومی دارد. بنابراین استفاده از روش ترکیبی PSO-GSA، می‌تواند مشکل کندی سرعت در تکرارهای آخر را برطرف سازد و همچنین با جلوگیری از گیر افتادن در نقاط کمینه‌ی محلی، به طور مؤثری به مقدار بهینه‌ی عمومی برسد. در الگوریتم PSO-GSA، تمام عوامل برای اولین بار به طور تصادفی مقداردهی اولیه می‌شوند و سپس نیروی گرانشی، ثابت گرانشی، نیروهای موجود بین عوامل و همچنین شتاب ذرات، مطابق با الگوریتم GSA محاسبه می‌شود. در نهایت، بعد از محاسبه شتاب و بهترین راه‌حل، سرعت و موقعیت تمام ذرات، مشابه با الگوریتم PSO و توسط روابط (۱۳) و (۱۴) به روزرسانی می‌شود [۳۰]:

مختلف بهبود یافته و ضعف‌های موجود در کارهای پیشین را تا حد امکان برطرف کرده است. در ادامه به بررسی هر یک از این مراحل پرداخته می‌شود.

۴-۱- پیش پردازش

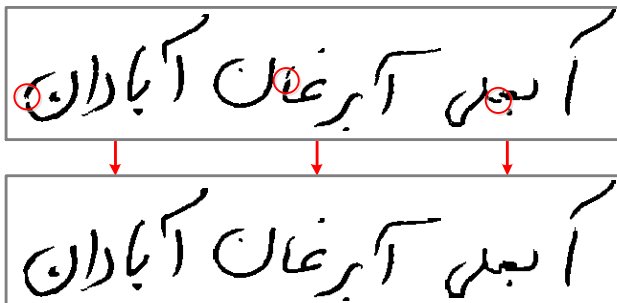
تصاویر ورودی برای اعمال به سیستم بازشناسی نوری کاراکتر (OCR)، باید ابتدا مورد پیش‌پردازش قرار بگیرند؛ تا عوامل ناخواسته‌ی آن‌ها حذف و برای مراحل بعد مناسب شوند. این مراحل شامل: باینری کردن تصویر، حذف نویز، اصلاح شکستگی‌های کلمات، حذف حاشیه‌های دور تصویر کلمه با استفاده از هیستوگرام افقی و عمودی و یکسان سازی اندازه‌ی تصاویر می‌باشد.

• باینری کردن تصویر

در این مرحله، تمامی تصاویر ورودی با استفاده از آستانه‌گذاری سراسری با روش اتسو^{۳۲} باینری شده و پس از آن تصویر مکمل (پیش‌زمینه سیاه و زیر کلمه سفید) به دست می‌آید.

• اصلاح محل شکستگی‌ها

از آنجا که سیستم بازشناسی پیشنهادی، براساس قطعه‌بندی زیر کلمات و سپس بازشناسی آن‌ها می‌باشد، اصلاح شکستگی‌های کلمات از اهمیت زیادی برخوردار است. شکستگی به نقاط جدا شده‌ی ناخواسته‌ی اطلاق می‌شود، که معمولاً فواصل آن‌ها در مقایسه با فواصل زیر کلمات کمتر است. برای اصلاح شکستگی‌ها، از تکنیک اتساع^{۳۳} با عضو ساختاری مستطیلی با ابعاد ۵×۳ و سپس فرسایش^{۳۴} با همین عضو ساختاری (عملیات بستن مورفولوژیکی) استفاده شده است [۳۳]. با این کار شکستگی‌های طولی کمتر از ۵ پیکسل و شکستگی‌های عرضی کمتر از ۳ پیکسل به یکدیگر متصل شده‌اند. شکل ۲، نمونه‌ای از شکستگی‌های ناخواسته زیر کلمات و اصلاح شده‌ی آن‌ها را با روش پیشنهادی نشان می‌دهد.



شکل ۲: اصلاح محل شکستگی کلمات دست‌نویس

• حذف نویز

برای حذف نویزهای ناخواسته، از فیلتر کردن اشیای دارای

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (19)$$

مجموعه عناصر x_i با مقادیر غیر صفر α_i بردارهای پشتیبان نامیده می‌شوند [۳۲]. در نهایت صفحه‌ای که به صورت بهینه داده‌ها را از هم جدا می‌کند، به صورت رابطه (۲۰) محاسبه می‌شود [۲۲،۳۲]:

$$f(x) = \text{sign}\left(\sum_{\text{support vector}} y_i \alpha_i (x_i^T x) + b\right) \quad (20)$$

به منظور اصلاح محدودیت‌های حاشیه‌ای، برای داده‌هایی که به صورت خطی قابل جداسازی نیستند، اصطلاحاً از حاشیه نرم^{۳۸} استفاده کرده و مجموعه‌ای به نام متغیرهای سست^{۳۹} (ϵ_i) تعریف می‌شود که در فرایند آموزش اجازه‌ی ورود به حاشیه را دارند. در نتیجه مسئله بهینه‌سازی به رابطه (۲۱) تبدیل می‌شود [۱۸،۳۲]:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \epsilon_i \quad (21)$$

پارامتر C ضریب جریمه است و برای برقراری مصالحه بین خطای طبقه‌بندی و بیشینه کردن حاشیه، استفاده می‌شود. این مسئله مشابه معادله (۱۷) حل می‌شود. البته در حل آن باید محدودیت‌های رابطه (۲۲) را برای ضرایب لاگرانژ در نظر گرفت [۳۲]:

$$0 \leq \alpha_i \leq C, \quad i = 1, \dots, N \quad (22)$$

الگوریتم SVM به جای تطبیق دادن منحنی‌های غیرخطی به داده‌ها، از تابع کرنل $K(x_i, x)$ برای نگاشت داده‌ها به فضایی با ابعاد بالاتر استفاده می‌کند. این فرآیند تبدیل توسط توابع کرنل مانند: خطی، چندجمله‌ای، تابع پایه‌ی شعاعی^{۳۰} (RBF) و سیگموئید^{۳۱} انجام می‌شود. RBF یک تابع مرسوم و قدرتمند در کاربردهای بازشناسی الگو است و به صورت رابطه (۲۳) تعریف می‌شود [۲۲]:

$$K(x_i, x) = \exp(-\gamma \|x - x_i\|^2) \quad (23)$$

که در آن γ پارامتر کرنل RBF می‌باشد. مشابه رابطه (۲۰) صفحه‌ای که به صورت بهینه داده‌ها را از هم جدا می‌کند، به صورت رابطه (۲۴) محاسبه می‌شود [۲۲،۳۲]:

$$f(x) = \text{sign}\left(\sum_{\text{support vector}} y_i \alpha_i K(x_i, x) + b\right) \quad (24)$$

۴- سیستم پیشنهادی

در این مطالعه، یک سامانه‌ی بازشناسی برون خط کلمات دست‌نویس فارسی متشکل از مراحل پیش‌پردازش، قطعه‌بندی، تهیه‌ی فرهنگ لغت تصویری، ایجاد پایگاه داده‌ی موقعیت زیر کلمات، استخراج ویژگی، طبقه‌بندی و پس‌پردازش، طراحی و پیاده‌سازی شده است. سیستم بازشناسی پیشنهادی در بخش‌های

- حروف "ف" و "ق" ابتدا و وسط را "ف" برچسب می‌زنیم.
- حروف "ر"، "ز" و "ژ" جدا و چسبیده را "ر" برچسب می‌زنیم.
- حروف "د"، "ذ" جدا و چسبیده را "د" برچسب می‌زنیم.
- حروف "ط" و "ظ" ابتدا، وسط و آخر را "ط" برچسب می‌زنیم.
- حروف "ص" و "ض" ابتدا، وسط و آخر را "ص" برچسب می‌زنیم.
- حروف "س" و "ش" ابتدا، وسط و آخر را "س" برچسب می‌زنیم.
- حروف "ح"، "ج"، "خ" و "چ" ابتدا، وسط و آخر را "ح" برچسب می‌زنیم.

- نقطه، ۲ نقطه و ۳ نقطه را به ترتیب 1D، 2D و 3D می‌نامیم.
- سرکش‌های جداشده‌ی "ک" و "گ" را ZB می‌نامیم.
- مدها را Mad و همزه‌ها را HZ می‌نامیم.
- برچسب حروف ابتدا، وسط و یا آخر باقیمانده، خود آن‌ها می‌باشد مانند: "ا"، "و"، "ل" و "ه".

شکل ۴ مراحل تهیه‌ی فرهنگ لغت تصویری زیرکلمات بدون علامت و برچسب زدن آن‌ها را نشان می‌دهد. همان‌گونه که مشاهده می‌شود، با اجرای الگوریتم بالا برای یک کلمه‌ی واحد ممکن است چند الگوی برچسب ایجاد شود؛ که این به دلیل تنوع نگارش کلمات دست‌نویس بوده و در مرحله ترکیب زیرکلمات بدون علامت، برای رسیدن به کلمه نهایی در نظر گرفته می‌شود. همچنین بعضی از کلمات مانند "آق قلا" از لحاظ تعداد زیرکلمات بدون علامت اصلی دارای دو شیوه‌ی نگارش می‌باشند (با زیرکلمه‌ی "قلا" دارای ۳ و با زیرکلمه‌ی "قل ا" دارای ۴ زیرکلمه‌ی بدون علامت اصلی می‌باشد) جهت حل این مسئله در پایگاه داده، برای این کلمات دو شیوه‌ی نگارش در نظر گرفته می‌شود (مانند "آق قلا - آق قل ا") به این ترتیب تعداد کلمات پایگاه داده از ۵۰۳ کلمه به ۵۳۱ کلمه افزایش می‌یابد. بعد از تهیه‌ی فرهنگ لغت تصویری، تعداد کلاس‌های آن برابر ۳۲۷ به-دست آمده، که تعداد ۵ کلاس مربوط به زیرکلمات بدون علامت فرعی (مد، تک نقطه، چند نقطه، همزه و سرکش) و ۳۲۲ کلاس مربوط به زیرکلمات بدون علامت اصلی می‌باشد.

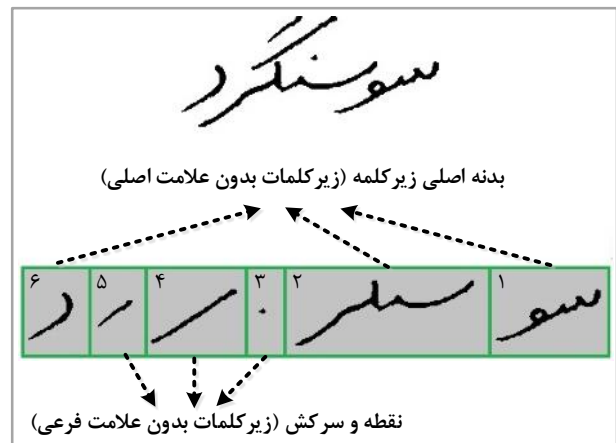
۴-۳- ایجاد پایگاه داده‌ی موقعیت زیرکلمات بدون علامت اصلی
ایجاد یک پایگاه داده از موقعیت زیرکلمات بدون علامت اصلی کمک می‌کند؛ تا برای بازشناسی یک زیرکلمه‌ی بدون علامت اصلی، براساس موقعیت آن زیرکلمه در کلمه‌ی اصلی تصمیم‌گیری شود. استفاده از این روش، باعث افزایش سرعت و بالا رفتن احتمال تشخیص درست در بازشناسی می‌گردد.

مساحت کمتر از ۱۰ پیکسل، استفاده شده است. به این ترتیب، اشیایی مانند نقاط، باقی‌مانده و نویزهای ریز حذف می‌شود.

• یکسان سازی اندازه‌ی تصاویر

تمامی تصاویر ورودی برای استفاده در مراحل بعدی سیستم بازشناسی، پس از باینری شدن، اصلاح شکستگی‌ها و حذف نویز، همگی به اندازه‌ی ۳۰۰×۳۰۰ پیکسل تبدیل شده‌اند.

۴-۲- قطعه‌بندی و تهیه‌ی فرهنگ لغت تصویری زیرکلمات بدون علامت
رویکرد به‌کار گرفته شده در بازشناسی کلمه، تشخیص تمامی اجزای متصل به هم در کلمه و سپس ترکیب آن‌ها با یکدیگر می‌باشد. این اجزای متصل به هم که زیرکلمات بدون علامت نام دارد؛ شامل زیرکلمات بدون علامت اصلی (مربوط به بدنه اصلی کلمه) و زیرکلمات بدون علامت فرعی (مربوط به نقطه، مد، همزه و سرکش) است (شکل ۳). فرهنگ لغت ایجاد شده باید تمام زیرکلمات بدون علامتی که در مرحله‌ی قطعه‌بندی از کلمه اصلی جدا شده است را شامل شود.



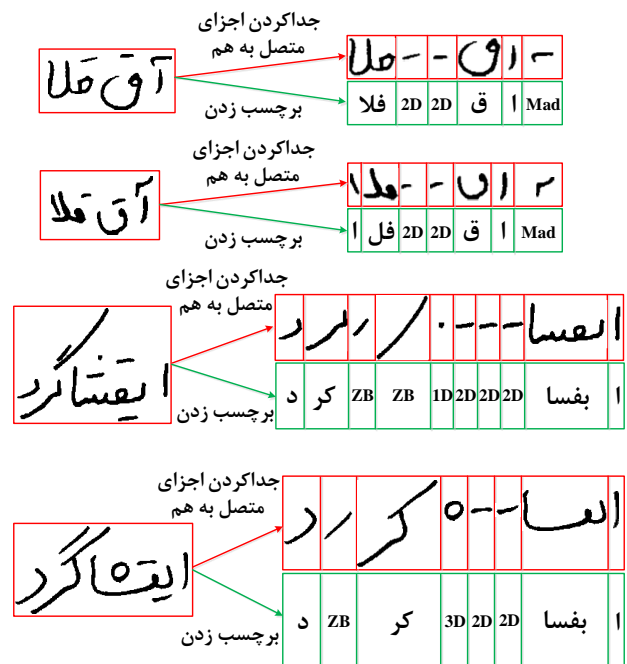
شکل ۳: اجزای متصل به هم کلمه شامل بدنه اصلی، نقطه و سرکش

برای تهیه فرهنگ لغت تصویری زیرکلمات بدون علامت، تمامی اجزای متصل به هم کلمات ورودی را جدا کرده و سپس آن‌ها را برچسب زده و در نهایت تمامی زیرکلمات بدون علامت با برچسب مشابه درون یک گروه قرار داده می‌شود. بعد از برچسب زدن اجزای متصل به هم تمامی کلمات دست‌نویس پایگاه داده، یک فرهنگ لغت تصویری به‌همراه برچسب متناظر آن‌ها ایجاد می‌شود؛ که می‌توان از آن برای آموزش سیستم بازشناسی استفاده کرد. در زیر قواعد برچسب زدن اجزای متصل به هم شرح داده می‌شود:

- حروف "ع" و "غ" ابتدا، وسط و آخر را "ع" برچسب می‌زنیم.
- حروف "ب"، "پ"، "ت" و "ث" ابتدا، وسط و آخر را "ب" برچسب می‌زنیم.
- حروف "ب"، "ن" و "ی" ابتدا و وسط را "ب" برچسب می‌زنیم.
- حروف "ک" و "گ" ابتدا، وسط و آخر را "ک" برچسب می‌زنیم.

جدول ۱: پایگاه داده موقعیت زیرکلمات

تعداد زیرکلمات اصلی	تعداد کلمات	موقعیت زیرکلمه	تعداد زیرکلمات بدون علامت
۱	۳۶	۱	۳۵
۲	۱۵۶	۱	۹۲
۳	۱۶۹	۲	۸۲
		۱	۶۴
۴	۱۲۱	۲	۶۰
		۳	۵۴
		۱	۳۹
۵	۳۸	۲	۳۹
		۳	۴۰
		۴	۴۰
		۱	۲۱
۶	۸	۲	۱۶
		۳	۱۲
		۴	۱۶
		۵	۱۴
		۱	۵
		۲	۵
۷	۳	۳	۷
		۴	۴
		۵	۶
		۶	۶
		۱	۲
		۲	۳
		۳	۳
مجموع کلمات	۵۲۱		



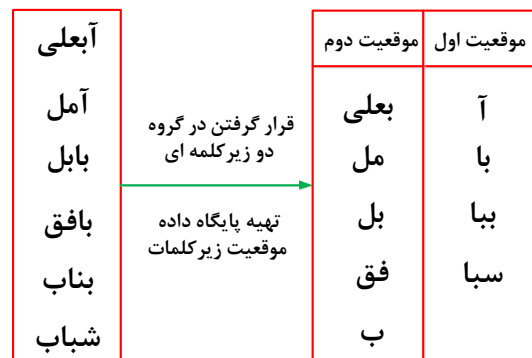
شکل ۴: زیرکلمات بدون علامت و برچسب زدن آن‌ها

جدول ۱ پایگاه داده تولید شده از موقعیت زیرکلمات را نشان می‌دهد. برای ایجاد این جدول تمامی ۵۳۱ کلمه ورودی از لحاظ تعداد زیرکلمات اصلی بررسی شده و در گروه‌های ۱ تا ۷ زیرکلمه اصلی دسته‌بندی می‌شود، که با این کار تعداد کلمات هر گروه مشخص می‌گردد. سپس برای هر موقعیت در هر گروه تعیین می‌شود که چند زیرکلمه بدون علامت غیرتکراری وجود دارد. به‌طور مثال، در پایگاه داده‌ی مورد استفاده، تعداد کلمات دو زیرکلمه‌ای ۱۵۶ مورد می‌باشد؛ که در موقعیت اول آن‌ها تعداد ۹۲ و در موقعیت دوم آن‌ها تعداد ۸۲ زیرکلمه‌ی بدون علامت غیرتکراری وجود دارد. شکل ۵، نحوه‌ی تهیه‌ی پایگاه داده موقعیت زیرکلمات را برای چند نمونه‌ی دو زیرکلمه‌ای نشان می‌دهد. همان‌گونه که واضح است با اعمال قواعد برچسب زدن اجزای متصل به هم، تعداد زیرکلمات بدون علامت غیرتکراری در هر موقعیت کاهش می‌یابد.

نکته قابل تأملی که مطرح می‌شود نحوه به‌روزرسانی این پایگاه داده (جدول ۱) در صورت اضافه شدن کلمات جدید است. در زیر الگوریتم پیشنهادی شرح داده می‌شود:

- ۱- اضافه شدن کلمه جدید به پایگاه داده.
- ۲- محاسبه‌ی تعداد زیرکلمات اصلی کلمه ورودی.
- ۳- مراجعه به ردیف مربوط به تعداد زیرکلمات اصلی و اضافه کردن یک واحد به تعداد کلمات آن.
- ۴- مراجعه به ردیف مربوط به موقعیت زیرکلمات.
- ۵- اگر زیرکلمه در موقعیت مربوطه وجود ندارد، تعداد زیرکلمات آن موقعیت یک واحد زیاد شده و زیرکلمه به پایگاه داده اضافه گردد.
- ۶- اگر زیرکلمه در موقعیت مربوطه وجود دارد، زیرکلمه به پایگاه داده اضافه نگردد.

مثلاً اگر کلمه‌ی جدید "تگزاس" باشد، بعد از برچسب زدن به



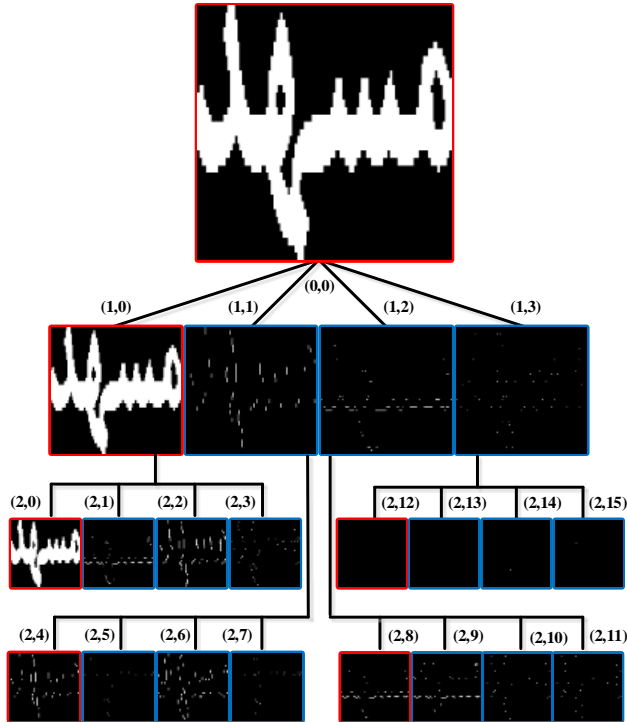
شکل ۵: نمونه‌ای از نحوه تهیه پایگاه داده موقعیت زیرکلمات

اطلاعات این مرحله برای استخراج ویژگی‌های مراحل ۲ و ۳ مورد استفاده قرار می‌گیرد. شکل ۶ درخت تجزیه‌ی بسته‌ی تبدیل موجک تا سطح دو موجک تا سطح دو

"بکراس" تبدیل می‌شود و تعداد زیرکلمات اصلی آن ۳ است که باید به ردیف سوم جدول ۱ مراجعه شود و به تعداد کلمات آن یک واحد اضافه گردد ($170=169+1$). برای زیرکلمه‌ی موقعیت اول یعنی "بکر" اگر در بین ۶۴ زیرکلمه موجود، وجود ندارد، آن را به جدول اضافه کرده و تعداد آن به‌روزرسانی می‌شود ($65=64+1$). و اگر در بین ۶۴ زیرکلمه موجود، وجود داشت، آن را به جدول اضافه نکرده و تعداد آن تغییر نمی‌کند. این روال برای زیرکلمات موقعیت دیگر نیز تکرار می‌شود.

۴-۴ استخراج ویژگی‌ها

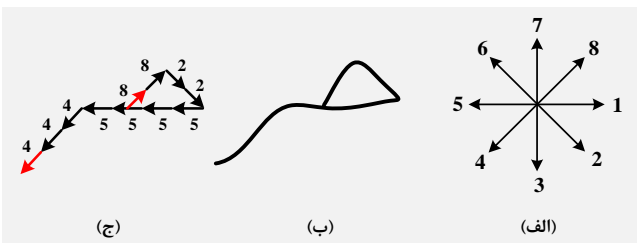
استخراج ویژگی یک مرحله‌ی مهم و حیاتی در سیستم‌های بازشناسی کاراکتر محسوب می‌شود؛ زیرا نتایج حاصل از این مرحله به‌طور مستقیم بر روی مرحله‌ی بازشناسی تأثیر می‌گذارد. در این بخش، یک روش استخراج ویژگی معرفی شده است که توصیف بسیار دقیقی از زیرکلمه دست‌نویس را فراهم می‌آورد. روش پیشنهادی شامل سه مرحله اصلی می‌باشد: در مرحله اول، ابتدا با استفاده از بسته‌ی تبدیل موجک، به حذف داده‌های اضافی تصاویر پرداخته می‌شود. در مرحله دوم، با معرفی کد زنجیره‌ای فریم دنباله مرکز جرم (CSFCC) به استخراج ویژگی پرداخته و در مرحله سوم با استخراج ویژگی‌های ساختاری بردار ویژگی نهایی ایجاد می‌شود.



شکل ۶: درخت تجزیه‌ی بسته‌ی تبدیل موجک تا سطح دو

مرحله دوم: کد زنجیره‌ای فریم (FCC) یک روش مؤثر استخراج ویژگی با در نظر گرفتن اطلاعات شکستگی‌ها، مرز و یا کانتور^{۳۸} اشیا است. برای استفاده از این ویژگی باید کانتور اشیا را بسته به نوع کاربرد به ۴، ۸، ۱۲ و یا ۳۶ جهت تجزیه کرد؛ که استفاده از تجزیه‌ی ۸ جهت مرسوم‌تر است [۳۳]. شکل ۷، نمونه‌ای از نحوه‌ی استخراج ویژگی کد زنجیره‌ای فریم ۸ جهت را برای زیرکلمه "حر" با ضخامت یک پیکسل، نشان می‌دهد. حاصل به ترتیب نگارش برابر "۸۸۲۲۵۵۵۵۴۴۴" می‌باشد.

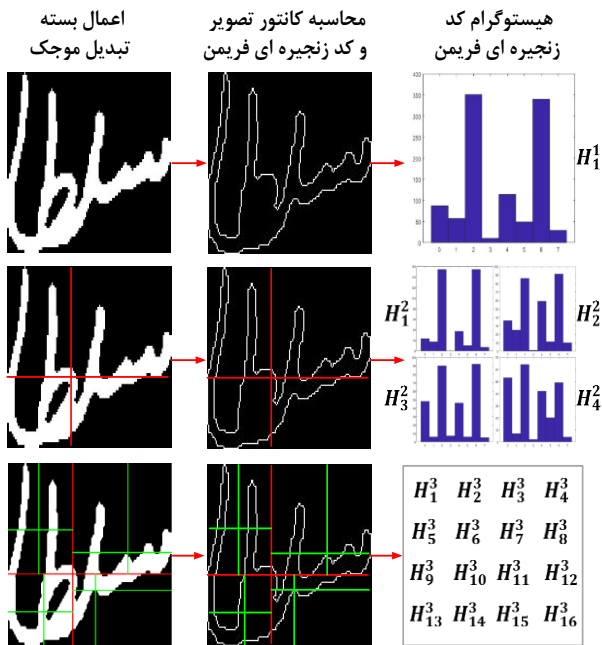
مرحله اول: روش بسته‌ی تبدیل موجک^{۳۵} (WPT)، حالت کلی‌تری از تجزیه‌ی موجک است. با استفاده از این تبدیل، می‌توان تصویر را به دو بخش اطلاعات تقریبی^{۳۶} و جزئی^{۳۷} تقسیم کرد. در تصاویر دودویی، اطلاعات تقریبی دربرگیرنده‌ی شکل کلی و اطلاعات جزئی عموماً شامل لبه‌ها و حفره‌های ریز می‌باشد. لبه‌های زیرکلمات نباید تیز و دارای تغییرات زیاد باشد زیرا این امر منجر به تولید کدهای زنجیره‌ای با تغییرات زیاد و نهایتاً تولید بردار ویژگی غیر مفید می‌گردد و دقت بازشناسی را تا حد زیادی کاهش می‌دهد. با به‌کارگیری بسته تبدیل موجک (یا اعمال سایر فیلترهای کاهش نویز و هموار سازی گوسی) می‌توان اطلاعات جزئی مربوط به زیرکلمات که عموماً مربوط به لبه‌های تصویر می‌باشند، حذف و لبه‌های زیرکلمات را تا حد مناسبی نرم کرد. انتخاب سطح تجزیه به‌صورت تجربی است و با آزمون و خطا مشخص می‌شود. در این مرحله، با استفاده از بسته‌ی تبدیل موجک با تابع تبدیل Haar به تجزیه‌ی تصویر زیرکلمه و استخراج اطلاعات از زیرباند (۲،۰) سطح دو پرداخته می‌شود. این زیرباند حاوی اطلاعات تقریبی و کلی از تصویر زیرکلمه است و منجر به حذف اطلاعات غیر ضروری، نرم‌تر شدن لبه‌ها و باقی ماندن اطلاعات مفید در زمینه‌ی بازشناسی تصاویر زیرکلمه می‌شود.



شکل ۷: (الف) کد زنجیره‌ای فریم ۸ جهت، (ب) زیرکلمه‌ی "حر" به ضخامت یک پیکسل (ج) تقریب زیرکلمه با کد زنجیره‌ای فریم

در این بخش، روش جدید و قدرتمند کد زنجیره‌ای فریم دنباله مرکز جرم (CSFCC) مطرح می‌گردد. به‌طور کلی این روش شامل

بردار ویژگی نهایی در این مرحله (FV_2)، که سایز آن برابر $178 = 2 \times 5 + 168$ است، مطابق معادله‌ی (۲۷) به دست می‌آید:

$$FV_2 = [H^T, C_1^2, C_1^3, C_2^3, C_3^3, C_4^3] \quad (27)$$


شکل ۸: محاسبه کد زنجیره‌ای فریمن دنباله مرکز جرم

مرحله سوم: با بررسی مقالات مرتبط با بازشناسی کلمات فارسی چاپی و دست‌نویس [۳، ۵، ۷، ۱۲] مشخص می‌گردد که بخش قابل توجهی از ویژگی‌های کلمات در نگارش فارسی، مربوط به ویژگی‌ها و مشخصه‌های ساختاری کلمه براساس قواعد رسم‌الخط می‌باشد. لذا در این تحقیق از ویژگی‌های ساختاری زیر جهت توصیف ساختار زیرکلمات استفاده شده است:

- مساحت زیرکلمه (S_1)
 - نسبت ارتفاع به عرض زیرکلمه قبل از یکسان سازی اندازه (S_2)
 - بیشینه هیستوگرام افقی و عمودی زیرکلمه (S_3, S_4)
 - بیشینه پیک‌های هیستوگرام افقی و عمودی زیرکلمه (S_5, S_6)
 - تعداد حفره‌های زیرکلمه (S_7)
 - مقدار شیب زیرکلمه با توجه به خط زمینه (S_8)
 - درصد پیکسل‌های بالا و پایین زیرکلمه نسبت به خط زمینه (S_9)
 - درصد پیکسل‌های چپ و راست زیرکلمه نسبت به خط وسط (S_{10})
- در این مرحله تعداد ۱۰ ویژگی ساختاری از تصویر زیرکلمه استخراج شده و مطابق فرمول (۲۸) با ویژگی‌های مرحله قبل ترکیب می‌شوند و بردار ویژگی نهایی (FV) با سایز $188 = 10 + 178$ را تشکیل می‌دهند:

$$FV = [FV_2, S_1, \dots, S_{10}] \quad (28)$$

اطلاعات کدهای زنجیری فریمن براساس مرکز جرم زیرکلمه به صورت دنباله‌ای و ترتیبی است. در این روش، تصویر زیرکلمه براساس اطلاعات مرکز جرم به بلوک‌های نامساوی تقسیم شده و سپس کانتور زیرکلمه در هر بلوک با استفاده از هیستوگرام کدهای زنجیری فریمن توصیف می‌شود. مزیت بلوک‌بندی تصویر با استفاده از مرکز جرم این است که در این روش تعداد بلوک‌های بدون اطلاعات (سیاه‌رنگ) در تصویر زیرکلمه به شدت کاهش یافته و بردار ویژگی بهینه‌تر می‌شود. زیرا این روش براساس تمرکز جوهر در هر ناحیه از تصویر (مرکز جرم)، زیرکلمه را بلوک‌بندی می‌کند و سپس کانتور تصویر توصیف می‌شود. مرکز جرم زیرکلمه با استفاده از فرمول (۲۵) محاسبه می‌گردد:

$$C_x = \frac{\sum_{x=1}^{x_{max}} x \sum_{y=1}^{y_{max}} b[x][y]}{\sum_{x=1}^{x_{max}} \sum_{y=1}^{y_{max}} b[x][y]} \quad C_y = \frac{\sum_{y=1}^{y_{max}} y \sum_{x=1}^{x_{max}} b[x][y]}{\sum_{x=1}^{x_{max}} \sum_{y=1}^{y_{max}} b[x][y]} \quad (25)$$

در معادله (۲۵)، x و y به ترتیب سطر و ستون تصویر، b مقدار پیکسل تصویر در نقطه مربوطه، x_{max} و y_{max} به ترتیب حداکثر سطر و ستون تصویر و C_x و C_y سطر و ستون مرکز جرم می‌باشد. استخراج ویژگی CSFCC شامل سه بخش اصلی می‌باشد که در ادامه تشریح می‌گردد.

۱. در بخش اول، کانتور تصویر زیرکلمه استخراج شده، سپس هیستوگرام کدهای زنجیره‌ای فریمن ۸ جهته بر روی این مرز توخالی محاسبه شده و بردار حاصل H_1^1 نامیده می‌شود. واضح است؛ به دلیل استفاده از هیستوگرام کد فریمن، نقطه شروع و همچنین دوران مرز اهمیتی ندارد. به دلیل استفاده از کدهای زنجیره‌ای فریمن ۸ جهته، بردار هیستوگرام حاصل ۸ مؤلفه دارد.

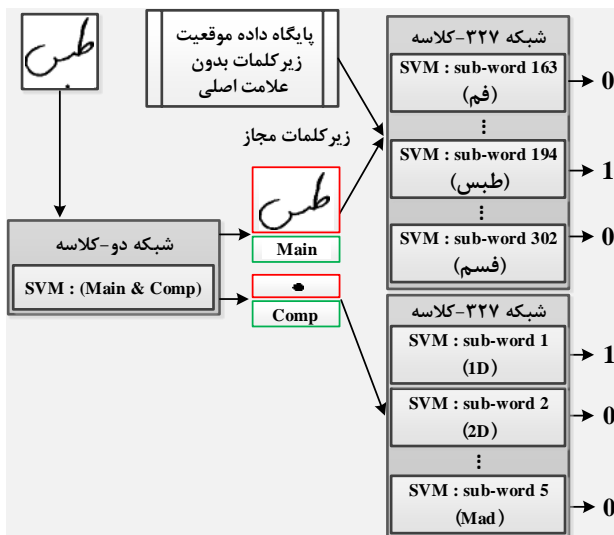
۲. در بخش دوم، مرکز جرم زیرکلمه ($C_1^2 = [C_x^2, C_y^2]$) محاسبه شده و تصویر از نقطه مرکز جرم به چهار بلوک نامساوی تقسیم می‌شود. سپس مشابه حالت قبل هیستوگرام کدهای FCC بر روی کانتور زیرکلمه، در هر بلوک به دست می‌آید. چهار بردار H_2^2, H_1^2, H_2^3 و H_4^2 در این بخش ایجاد می‌شود.

۳. در بخش سوم، دوباره مرکز جرم هر ۴ بلوک بخش قبلی ($C_4^3, C_3^3, C_2^3, C_1^3$) محاسبه شده و تصویر از نقاط مرکز جرم تقسیم می‌شود. در نتیجه ۱۶ بلوک نامساوی به دست می‌آید و با محاسبه هیستوگرام کدهای FCC در هر بلوک، بردارهای $H_1^3, H_2^3, \dots, H_{16}^3$ حاصل می‌شود (شکل ۸). از کنار هم قرار دادن این هیستوگرام‌ها بردار هیستوگرام کلی H^T مطابق معادله (۲۶) به دست می‌آید:

$$H^T = [H_1^1, H_1^2, \dots, H_4^2, H_1^3, \dots, H_{16}^3] \quad (26)$$

سایز بردار H^T برابر $168 = 8 + 4 \times 8 + 16 \times 8$ است. اطلاعات سطر و ستون ۵ نقطه مرکز جرم نیز به این بردار اضافه شده و

داده‌ی موقعیت زیرکلمات بدون علامت اصلی استفاده کرد. به همین منظور، یک شبکه SVM دو کلاسه ایجاد شده است؛ که خروجی آن برای تمامی زیرکلمات بدون علامت فرعی (نقطه، مد، همزه و سرکش) صفر و برای تمامی زیرکلمات بدون علامت اصلی یک است. این شبکه قابلیت تفکیک نوع زیرکلمه را ندارد و فقط فرعی و اصلی آن را مشخص می‌کند. بنابراین نیاز به یک شبکه دیگر وجود دارد که تمامی ۳۲۷ زیرکلمه ورودی را تشخیص دهد. هدف از ایجاد پایگاه داده‌ی موقعیت زیرکلمات بدون علامت اصلی (جدول ۱)، بازشناسی زیرکلمه براساس موقعیت آن در کلمه ورودی بود. در نتیجه به جای استفاده از یک طبقه‌بند بزرگ با تعداد بسیار زیاد خروجی، از چندین طبقه‌بند کوچک با تنها یک خروجی استفاده گردیده است. با این روش حجم شبکه به شدت کاهش یافته و فرآیند آموزش شبکه سرعت می‌گیرد. همچنین می‌توان با استفاده از اطلاعات پایگاه داده موقعیت زیرکلمات، تنها آن دسته از طبقه‌بندهایی را در فرآیند بازشناسی انتخاب نمود؛ که با توجه به موقعیت زیرکلمات آن‌ها در کلمه اصلی قابل قبول هستند. شکل ۹ به کارگیری ۳۲۷ طبقه‌بند SVM و رویکرد یک در برابر همه را نشان می‌دهد. همان‌گونه که مشخص است؛ از دید هر SVM، اگر زیرکلمه‌ی ورودی با کلاس خودش مطابق باشد؛ خروجی آن یک و در غیر این صورت برای سایر کلاس‌ها خروجی صفر می‌شود.



شکل ۹: به کارگیری چندین طبقه‌بند SVM و رویکرد یک در برابر همه

۴-۶- ترکیب زیرکلمات و پس‌پردازش

جهت اصلاح برخی از اشتباهات اتفاق افتاده و در نهایت بهبود نرخ بازشناسی، استفاده از اطلاعات معنایی و شکلی در یک سیستم OCR ضروری است. اغلب روش‌های پس‌پردازش براساس اطلاعاتی مانند فرهنگ لغت، اطلاعات آماری، اطلاعات دستوری و معنایی،

۴-۵- آموزش سیستم پیشنهادی

آموزش سیستم بازشناسی پیشنهادی شامل دو مرحله اصلی می‌باشد. در مرحله اول طبقه‌بند SVM توسط الگوریتم PSOGSA به گونه‌ای بهینه می‌گردد که خطای طبقه‌بندی کمینه شود. مرحله بعدی شامل آموزش دو طبقه‌بند متفاوت می‌باشد. طبقه‌بند اول، یک شبکه‌ی دو کلاسه است که زیرکلمات اصلی و فرعی را از هم جدا می‌کند و طبقه‌بند دوم وظیفه‌ی تشخیص زیرکلمه را بر عهده دارد. در ادامه مراحل مذکور شرح داده شده است.

۴-۵-۱- الگوریتم PSOGSA-SVM

همان‌گونه که پیش‌تر عنوان شد، در این پژوهش تابع کرنل الگوریتم SVM، از نوع RBF انتخاب گردیده است. در الگوریتم SVM انتخاب مناسب پارامتر کرنل (γ) و ضریب جریمه (C) اهمیت بسیار زیادی دارد؛ به گونه‌ای که تغییرات اندک در مقادیر این پارامترها باعث تغییرات زیادی در خطای طبقه‌بندی می‌شود [۳۴]. در روش پیشنهادی، الگوریتم SVM با استفاده از الگوریتم PSOGSA به گونه‌ای بهینه می‌گردد؛ که خطای کلاس‌بندی کمینه شود. این کار با بهینه کردن پارامترهای کنترلی الگوریتم SVM شامل ضریب جریمه (C) و پارامتر کرنل (γ) انجام می‌شود. این پارامترها در کنار هم بردار موقعیت ذرات را تشکیل می‌دهند و الگوریتم PSOGSA با کمینه کردن تابع برازندگی و به‌روزرسانی بردار موقعیت ذرات، الگوریتم SVM را بهینه می‌کند. تابع خطای طبقه‌بندی مطابق رابطه (۲۹) به‌عنوان تابع برازندگی در نظر گرفته می‌شود:

$$f(C, \gamma) = 100 - SVM(C, \gamma)_{Acc} \quad (29)$$

$SVM(C, \gamma)_{Acc}$ درصد دقت طبقه‌بندی SVM با پارامترهای (C, γ) و $f(C, \gamma)$ درصد خطای طبقه‌بندی می‌باشد. در زیر مراحل الگوریتم PSOGSA-SVM به‌صورت کامل تشریح شده است:

- ۱- تعیین پارامترهای الگوریتم PSOGSA $(\alpha, G_0, C_1, C_2, \omega)$.
- ۲- تولید جمعیت اولیه ذرات ($nPop$).
- ۳- ارزیابی تابع برازندگی $f(C, \gamma)$ برای تمام ذرات.
- ۴- به‌روزرسانی G و $gbest$ برای تمام ذرات.
- ۵- محاسبه‌ی جرم، نیرو و شتاب تمام ذرات.
- ۶- به‌روزرسانی بردار سرعت و موقعیت ذرات (C, γ).
- ۷- رفتن به مرحله ۳ تا زمان رسیدن به شرط توقف (تعداد تکرار).
- ۸- در پایان تکرارها، مقدار بهینه پارامترهای (C, γ) ذخیره شود.

۴-۵-۲- استفاده از چندین طبقه‌بند SVM

در ابتدای فرآیند بازشناسی باید اصلی و فرعی بودن تمام زیرکلمات کلمه‌ی ورودی تعیین شود؛ تا بتوان از اطلاعات پایگاه-

مقدار ۸۵ درصد از تصاویر هر کلمه برای آموزش (حدود ۲۵ تا ۳۴ تصویر از هر کلمه) و ۱۵ درصد باقیمانده (حدود ۵ تا ۶ تصویر از هر کلمه) برای آزمایش سیستم پیشنهادی در نظر گرفته شده است. بعد از تهیه‌ی فرهنگ لغت تصویری زیرکلمات بدون علامت از روی تصاویر آموزش، تعداد ۳۲۷ کلاس زیرکلمه بدون علامت اصلی و فرعی ایجاد می‌شود؛ که از آن‌ها برای آموزش و اعتبارسنجی^{۳۹} سیستم پیشنهادی استفاده می‌گردد. پایگاه داده‌ی استفاده شده در این پژوهش، یک پایگاه داده‌ی قدرتمند در زمینه‌ی ارزیابی سیستم بازشناسی کلمات دست‌نویس فارسی بوده و در مقالات متعددی مورد استفاده قرار گرفته است. در ادامه توضیح مختصری از چند مقاله‌ی نمونه، که از پایگاه داده ایرانشهر استفاده کرده‌اند، ارائه کرده و نتایج سیستم پیشنهادی با آن‌ها مقایسه می‌گردد.

دهقان و همکاران در [۳۷] با استفاده از مدل مخفی مارکف (HMM) گسسته و بردار ویژگی خودسازمانده (SOFM)^{۴۰} به منظور نرم کردن توزیع احتمال مشاهده شده در فرایند آموزش HMM، به بازشناسی کلمات دست‌نویس فارسی پرداخته‌اند. برومند نیا و همکاران در [۳۸] با بهره‌گیری از بسته‌ی تبدیل موجک M_band قطبی به استخراج ویژگی‌های مستقل از چرخش و مقیاس از کلمات فارسی و طبقه‌بندی آن‌ها با الگوریتم ماهالانویس^{۴۱} پرداخته‌اند. یونسی و کبیر در [۳۹] به جای استفاده از یک طبقه‌بند پیچیده با تعداد زیادی خروجی، چندین طبقه‌بند از شبکه عصبی بازگشتی^{۴۲} (RNN) با خروجی‌های باینری در نظر گرفتند و فرایند آموزش شبکه را تسهیل کردند. قادیکلایی و همکاران در [۶] از یک الگوریتم قطعه‌بندی تبدیل کلمات به زیرکلمات شامل نقطه، سرکش و مد و چندین طبقه‌بند (RNN)، برای بازشناسی کلمات دست‌نویس فارسی استفاده کرده‌اند.

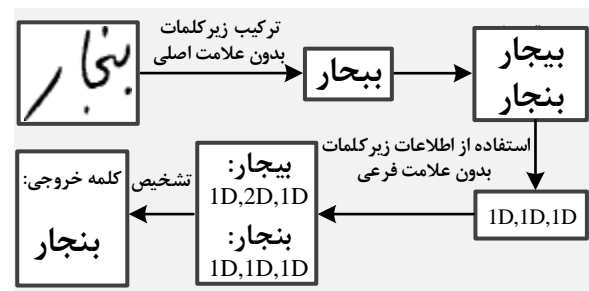
۶- ارزیابی الگوریتم پیشنهادی

در بخش نخست ارزیابی، مقدار بهینه‌ی پارامترهای طبقه‌بند SVM با استفاده از الگوریتم PSO^{۴۳} برابر $C = 34.23$ و $\gamma = 3.12$ به دست آمده است. همچنین یک آنالیز حساسیت به‌منظور نشان دادن اهمیت این پارامترها در عملکرد طبقه‌بند SVM صورت گرفته که نتایج آن در جدول ۲ آورده شده است. با توجه به نتایج به‌دست آمده از جدول ۲، مشاهده می‌شود؛ که استفاده از الگوریتم PSO^{۴۳} در تعیین مقدار بهینه‌ی پارامترهای (C) و (γ) بسیار مؤثر بوده و تغییر اندک در مقدار این پارامترها نتیجه‌ی قابل توجهی در دقت نهایی سیستم بازشناسی خواهد داشت.

در جدول ۳ نتایج بازشناسی سیستم پیشنهادی به تفکیک استفاده از ویژگی‌ها آمده است. با توجه به نتایج به‌دست آمده، استفاده

نتایج حاصل از بازشناسی را بهبود می‌بخشند [۴]. مسئله کلیدی در مرحله پس‌پردازش، چگونگی استفاده از اطلاعات معنایی جهت انتخاب یک کلمه‌ی قابل قبول از میان کلمات متعدد کاندید شده به‌عنوان نتیجه‌ی نهایی می‌باشد [۳۵].

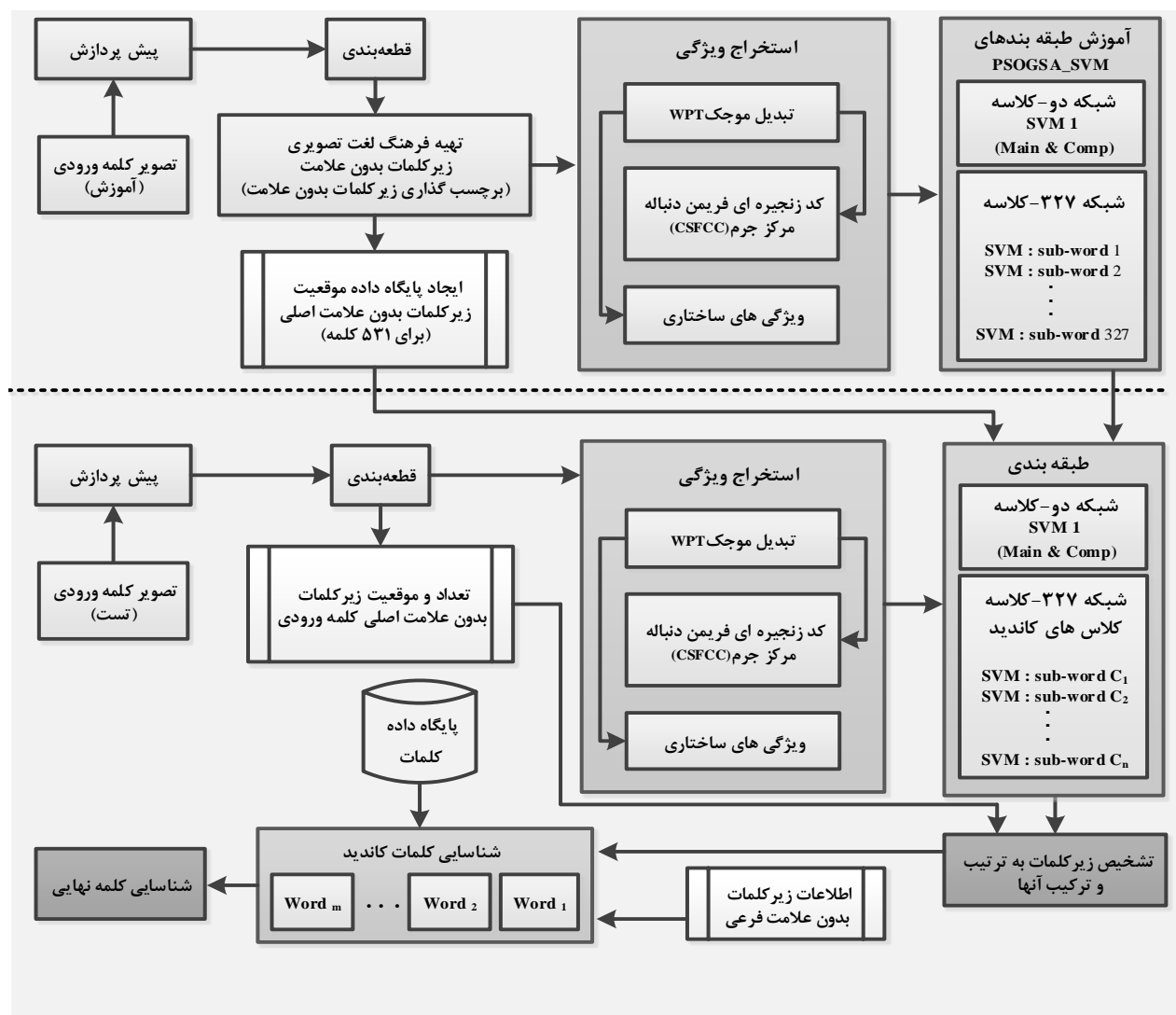
در روش پیشنهادی، بعد از تشخیص زیرکلمات به‌صورت ترتیبی، آن‌ها را با یکدیگر ترکیب کرده، تا کلمه موردنظر ورودی تشکیل شود و سپس براساس تعداد زیرکلمات بدون علامت اصلی، نزدیک‌ترین کلمه‌ی پایگاه داده، به کلمه تشخیص داده شده، به‌عنوان خروجی گزارش می‌شود. از آنجا که فرآیند تشخیص براساس زیرکلمات بدون علامت اصلی صورت می‌گیرد؛ در موارد محدودی چند کلمه از پایگاه داده تشخیص داده می‌شود و تصمیم‌گیری بین آن‌ها نیازمند اطلاعات زیرکلمات بدون علامت فرعی می‌باشد که در تمام این موارد با استفاده از دنباله‌ی زیرکلمات بدون علامت فرعی و بدون توجه به موقعیت آن‌ها می‌توان کلمه نهایی را تشخیص داد. البته در صورتی که پایگاه داده بزرگ‌تر و کلماتی در پایگاه داده وجود داشته باشد که تفاوت آن‌ها در موقعیت زیرکلمات بدون علامت فرعی باشد (مانند "باب" و "ناب")، می‌توان با در نظر قرار دادن اطلاعات موقعیت مکانی زیرکلمات بدون علامت فرعی، نسبت به بدنه اصلی و یا حتی خط زمینه، کلمه‌ی نهایی را تشخیص داد. شکل ۱۰، مراحل ترکیب زیرکلمات و تشخیص کلمه‌ی خروجی را نشان می‌دهد. همچنین روندنمای سیستم پیشنهادی در شکل ۱۱ به‌صورت کامل نشان داده شده است.



شکل ۱۰: مراحل ترکیب زیرکلمات و پس‌پردازش

۵- پایگاه داده

در این مقاله، از پایگاه داده پرکاربرد و شناخته شده Iranshahr استفاده شده است؛ که شامل ۵۰۳ نام شهر ایران با دست‌خط‌های متنوع است و تمامی تصاویر در کیفیت ۹۶ نقطه در اینچ و در فرم سطح خاکستری اسکن شده‌اند [۶،۳۶]. برای هر نام شهر تعداد حدود ۳۰ تا ۴۰ نمونه‌ی مختلف تصویر وجود دارد و در مجموع بیش از ۱۷۰۰۰ تصویر را شامل می‌شود. در بخش آموزش و آزمایش، تصاویر به‌صورت تصادفی به دو گروه تقسیم شده است.



شکل ۱۱: روندنمای سیستم بازشناسی پیشنهادی

جدول ۲: نتایج سیستم بازشناسی با تغییر پارامترهای C و γ

پارامتر C	پارامتر γ	نرخ بازشناسی
۳۴,۲۳	۰,۰۱	٪۴۳
۳۴,۲۳	۰,۱	٪۵۸
۳۴,۲۳	۰,۵	٪۶۰
۳۴,۲۳	۱	٪۷۰
۳۴,۲۳	۳,۱۲	٪۸۹,۲۹
۳۴,۲۳	۴	٪۸۴
۲۰	۳,۱۲	٪۸۳
۱۰	۳,۱۲	٪۸۰

در جدول ۴، نرخ بازشناسی روش پیشنهادی با سایر کارهای تشریح شده، مقایسه گردیده است. نتایج به دست آمده، کارایی روش پیشنهادی را در مقایسه با روش‌های موجود نشان می‌دهد.

هم‌زمان از ویژگی جدید و قدرتمند کد زنجیره‌ای فریمین دنباله مرکز جرم (CSFCC) همراه با پیش‌پردازش بسته‌ی تبدیل موجک و ویژگی‌های ساختاری، موجب بالا رفتن نرخ بازشناسی شده است. همچنین استفاده از ویژگی CSFCC به‌تنهایی و یا در کنار ویژگی‌های ساختاری نیز عملکرد قابل قبولی را از خود نشان داده است؛ که این نشان‌دهنده‌ی قدرت ویژگی CSFCC در توصیف مؤثر یک کلمه‌ی دست‌نویس می‌باشد. از نتایج به دست آمده واضح است که استفاده از ویژگی‌های ساختاری به‌تنهایی نمی‌تواند در بازشناسی کلمات دست‌نویس فارسی مؤثر باشد. نکته قابل توجه عدم بهبود مؤثر دقت ویژگی‌های ساختاری، در استفاده از پیش‌پردازش بسته‌ی تبدیل موجک است. که علت این امر مستقل بودن ویژگی‌های ساختاری از کانتور کلمات است. در مورد ویژگی CSFCC، به دلیل وابستگی به کانتور، استفاده از WPT، به‌عنوان یک راه مؤثر در جهت افزایش نرخ بازشناسی قابل توجه است.

روش پیشنهادی را به منظور افزایش سرعت شناسایی در کاربردهای بلادرنگ، بر روی GPU پیاده‌سازی کرد. همچنین می‌توان از شبکه عصبی پیچشی (CNN)، به‌عنوان طبقه‌بند در فرآیند بازشناسی استفاده نمود. برای کاهش زمان بازشناسی و افزایش دقت، می‌توان مرحله‌ی انتخاب ویژگی را در سیستم بازشناسی پیشنهادی در نظر گرفت و از الگوریتم‌های مبتنی بر جستجوی تصادفی مانند: الگوریتم ژنتیک و یا بهینه‌سازی ازدحام ذرات، جهت انتخاب ویژگی‌های مناسب استفاده کرد. همچنین می‌توان به‌منظور کاهش طول کد زنجیره فریمن به‌جای استفاده از تبدیل موجک از سایر روش‌ها مانند فیلترهای کاهش نویز و هموار سازی گوسی استفاده و آن را ارزیابی نمود.

مراجع

- [۱] A. A. Aburas and S. M. Rehie, "OffLine Omni-Style Handwriting Arabic Character Recognition System Based on Wavelet Compression", Arab Research Institute in Sciences & Engineering, vol.3, No.4, pp. 123-135, 2007.
- [۲] S. Nasrollahi and A. Ebrahimi, "Printed Persian Subword Recognition Using Wavelet Packet Descriptors", Journal of Engineering, Vol.2013, 2013.
- [۳] امید هاشمی قوچانی و علیرضا سیدین، "شناسایی برون خط کلمات دست‌نویس فارسی با تاکید بر تشخیص نام چند شهر"، هفتمین کنفرانس ماشین بینایی و پردازش تصویر ایران، آبان ۱۳۹۰.
- [۴] N. Arica and T. Yamin-Vural Fatos, "An overview of character recognition based focused on offline handwriting", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol.31, No.2, 2001.
- [۵] M. Zand, A. N. Nilchi and S. A. Monadjemi, "Recognition-based Segmentation in Persian Character Recognition", World Academy of Scienci, Engineering and Technology, Vol.38, pp. 183-187, 2008.
- [۶] M. F. Y. Ghadikolaie, E. Kabir and F. Razzazi, "Sub-word Based Offline Handwritten Farsi Word Recognition Using Recurrent Neural Network", ETRI Journal, Vol.38, N.4, pp. 703-713, 2016.
- [۷] M. Omidyeganeh, K. Nayebe, R. Azmi and A. Javadtalab, "A new segmentation technique for multi font farsi/arabic texts", IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.2, pp. 757-760, 2005.
- [۸] M. Omidyeganeh, R. Azmi, K. Nayebe, and A. Javadtalab, "A new method to improve multi font Farsi/arabic character segmentation results: using extra classes of some character combinations", International Conference on Multimedia Modeling, pp. 670-679, 2007.
- [۹] G. Katiyar and S. Mehruz, "A hybrid recognition system for offline handwritten characters" SpringerPlus, Vol. 5, No.1, 2016.
- [۱۰] H. Soltanzadeh and M. Rahmati, "Recognition of Persian handwritten digit using image profiles of multiple orientations", Pattern Recognition Letters, Vol.25, No.14, pp. 1569-1576, 2004.
- [۱۱] G. Vamvakas, B. Gatos and S.J. Perantonis, "Handwritten character recognition through two stage foreground sub-sampling", Pattern Recognition, Vol.43, No.8, pp. 2807-2816, 2010.
- [۱۲] K. Bagheri Noaparast and A. Broumandnia, "Persian handwritten word recognition using Zernike and fourier-mellin moments", 5th International Conference Sciences of Electronic, Technologies of Information and Telecommunications, 2009.
- [۱۳] H. Karimi, A. Esfahanimehr, M. Moseleh, F. M. Jadval ghadam, S. Salehpour, and O. Medhati, "Persian handwritten digit recognition using ensemble classifiers", Procedia Computer Science, Vol.73, pp. 416-425, 2015.
- [۱۴] P. Zhang, T. D. Bui, and C. Y. Suen, "A novel cascade ensemble classifier system with a high recognition performance on

جدول ۳: نتایج روش پیشنهادی به تفکیک ویژگی‌های مختلف

ویژگی‌های استفاده شده	نرخ بازشناسی
ساختاری بدون پیش‌پردازش WPT	۴۲٫۷۲٪
ساختاری با پیش‌پردازش WPT	۴۳٫۲۶٪
CSFCC بدون پیش‌پردازش WPT	۷۸٫۲۲٪
CSFCC با پیش‌پردازش WPT	۸۷٫۰۱۱٪
ساختاری و CSFCC بدون پیش‌پردازش WPT	۸۰٫۴۱٪
ساختاری و CSFCC با پیش‌پردازش WPT	۸۹٫۲۹٪

جدول ۴: مقایسه روش پیشنهادی با سایر مقالات

نویسندگان	نرخ بازشناسی
دهقان و همکاران [۳۷]	۶۳٪
برومندنیا و همکاران [۳۸]	۷۵٫۵٪
یونسی و کبیر [۳۹]	۸۳٫۹٪
قادیلایی و همکاران [۶]	۸۴٫۳٪
روش پیشنهادی	۸۹٫۲۹٪

۷- نتیجه‌گیری و کارهای آتی

در این مقاله به ارائه روشی جدید برای بازشناسی برون خط کلمات فارسی دستنویس پرداخته شد. رویکرد ارائه شده جهت بازشناسی، تشخیص تمامی اجزای متصل بهم در کلمه و سپس ترکیب آنها با یکدیگر می‌باشد. به همین منظور یک فرهنگ لغت تصویری زیر کلمات بدون علامت تهیه گردید. هدف اصلی در این تحقیق، تمرکز بر روی چگونگی استخراج ویژگی‌های مؤثر جهت افزایش نرخ بازشناسی بوده است. بدین منظور در روش پیشنهادی ابتدا با استفاده از یک مرحله پیش‌پردازش مبتنی بر اعمال بسته‌ی تبدیل موجک، اطلاعات غیرضروری داده‌های ورودی حذف و تصاویر برای مرحله‌ی استخراج ویژگی آماده شدند. در مرحله استخراج ویژگی، از ویژگی جدید و قدرتمند کد زنجیره‌ای فریمن دنباله مرکز جرم (CSFCC) به همراه ویژگی‌های ساختاری استفاده گردید که این روش استخراج ویژگی، موجب بالا رفتن نرخ بازشناسی شد. در بخش طبقه‌بندی از چندین طبقه‌بند SVM دو کلاسه بهبود یافته با الگوریتم تکاملی PSO GSA استفاده شده است. همچنین با ایجاد یک پایگاه داده از موقعیت زیر کلمات، تنها آن دسته از طبقه‌بندهایی در فرآیند شناسایی انتخاب شدند؛ که با توجه به موقعیت زیر کلمات آنها در کلمه اصلی قابل قبول بودند. این کار باعث، کاهش فضای جستجو و بالا رفتن احتمال تشخیص درست گردید. دقت بازشناسی به دست آمده در روش پیشنهادی، ۸۹٪ در بخش آزمایش می‌باشد؛ که نشان‌دهنده توانمندی روش پیشنهادی و بهبود نتایج در مقایسه با دیگر روش‌های موجود است. در کارهای آینده در راستای شناسایی کلمات دست‌نویس می‌توان،

- [۳۴] Q. H. Mao, H. W. Ma, and X. H. Zhang, "SVM Classification Model Parameters Optimized by Improved", *Advanced Materials Research*, Vols.889-890, pp. 617-621, 2014.
- [۳۵] Z. Hu, J. Lin, and L. Wu, "Research on OCR Post-processing Applications for Handwritten Recognition Based on Analysis of Scientific Materials", In *Advances in Computer Science, Intelligent System and Environment*, pp. 131-135, 2011.
- [۳۶] E. Bayesteh, A. Ahmadifard and H. Khosravi, "A Lexicon Reduction Method Based on Clustering Word Images in Offline Farsi Handwritten Word Recognition Systems", *Machine Vision and Image Processing*, pp. 1-5, 2011.
- [۳۷] M. Dehghan, K. Faez, M. Ahmadi, and M. Shridhar, "Handwritten Farsi (Arabic) Word Recognition: A Holistic Approach Using Discrete HMM", *Pattern Recognition*, Vol.34, No.5, pp. 1057-1065, 2001.
- [۳۸] A. Broumandnia, J. Shanbehzadeh and M. R. Varnoosfaderani, "Persian/Arabic Handwritten Word Recognition Using M-Band Packet Wavelet Transform", *Image and Vision Computing*, Vol.26, No.6, pp. 829-842, 2008.
- [۳۹] M. F. Younessy and E. Kabir, "A New Classifier Based on Recurrent Neural Network Using Multiple Binary-Output Networks", *IOSR Journal of Computer Engineering*, Vol.17, No.3, pp. 63-69, 2015.
- handwritten digits", *Pattern Recognition*, Vol.40, No.12, pp. 3415-3429, 2007.
- [۱۵] A. Mowlaei, K. Faez, and A. T. Haghghat, "Feature Extraction with Wavelet Transform for Recognition of Isolated Handwritten Farsi /Arabic Characters and Numerals", *Digital Signal Processing*, Vol.2, pp. 923-926, 2002.
- [۱۶] R. Sheikhpour, M. A. Sarram, S. Gharaghani and M. A. Z. Chahooki, "A Survey on semi-supervised feature selection methods", *Pattern Recognition*, Vol.64, pp. 141-158, 2017.
- [۱۷] R. Azmi, B. Pishgoo, N. Norozi, M. Koohzadi and F. Baesi, "A hybrid GA and SA algorithms for feature selection in recognition of handprinted Farsi characters", *IEEE International Conference on Intelligent Computing and Intelligent Systems*, Vol.3, pp. 384-387, 2010.
- [۱۸] N. Shanthi, and K. Duraiswamy, "A novel SVM-based handwritten Tamil character recognition system", *Pattern Analysis & Applications*, Vol.13, No. 2, pp. 173-180, 2009.
- [۱۹] C. J. Burges, "A tutorial on support vector machines for pattern recognition", *Data Mining Knowledge Discovery*, Vol.2, No.2, pp. 121-167, 1998.
- [۲۰] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A training algorithm for optimal margin classifiers", In *Proceedings of the fifth annual workshop on Computational learning theory*, 1992.
- [۲۱] Y. Wu, F. Yin and C. Liu, "Improving handwritten Chinese text recognition using neural network language models and convolutional neural network shape models", *Pattern Recognition*, Vol. 65, pp. 251-264, 2017.
- [۲۲] M. Elleuch, R. Maalej and M. Kherallah, "A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition", *Procedia Computer Science*, vol.80, pp. 1712-1723, 2016.
- [۲۳] M. Dehghan and K.Faez, "Handwritten Farsi Character Recognition Using Evolutionary Fuzzy Clustering", *Signal Processing Conference*, 1998.
- [۲۴] J. H. AlKhateeb, J. Jiang, J. Ren, F. Khelifi, S. S. Ipson, "Multiclass Classification of Unconstrained Handwritten Arabic Words Using Machine Learning Approaches", *The Open Signal Processing Journal*, vol.2, No.1, pp. 21-28, 2009.
- [۲۵] A. Ebrahimi and E. Kabir, "A pictorial dictionary for printed farsi subwords", *Pattern Recognition Letters*, Vol.29, No.5, pp. 656-663, 2008.

زیر نویس ها:

- ¹ Segmentation
² Feature Extraction
³ Feature Selection
⁴ Classification
⁵ Classification and Recognition
⁶ Post processing
⁷ Local Chain Code
⁸ Vertical and Horizontal Projections
⁹ Wavelet
¹⁰ Simulated Annealing
¹¹ Template Matching
¹² Support Vector Machines
¹³ Neural Networks
¹⁴ Convolutional Neural Networks
¹⁵ Recurrent Neural Networks
¹⁶ K-Nearest-Neighbours
¹⁷ K-means
¹⁸ Hidden Markov Model
¹⁹ Dropout
²⁰ Extreme Learning Machine
²¹ Fuzzy c-Means clustering
²² Evolutionary Strategies
²³ Centroid Sequence Freeman Chain Code
²⁴ Particle Swarm Optimization
²⁵ Gravitational Search Algorithm
²⁶ Margin
²⁷ Dual Lagrangian
²⁹ Soft Margin
³⁰ Slack Variables
³⁰ Radial Basis Function
³¹ Sigmoid
³² Otsu
³⁴ Dilation
³⁵ Erosion
³⁵ Wavelet Packet Transform
³⁶ approximate
³⁷ detail
³⁸ contour
⁴⁰ Validation
⁴¹ self-organizing feature map
⁴¹ Mahalanobis
⁴² Recurrent Neural Network

- [۲۶] مسعود فرکی، ومازیار پالهنگ، "بازشناسی برخط حروف فارسی بر پایه مدل مخفی مارکوف"، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۰، شماره ۱، صفحه ۳۴-۲۳، ۱۳۸۹.
- [۲۷] محمد نحوی، مهدی رفیعی، رضا ابراهیم پور و احسان اله کبیر، "ترکیب طبقه‌بندهای دوکلاسی برای بازشناسی ارقام دست‌نویس فارسی"، *شانزدهمین کنفرانس مهندسی برق ایران*، تهران، دانشگاه تربیت مدرس، ۱۳۸۷.
- [۲۸] Z. Tamen, H. Drias and D. Boughaci, "An efficient multiple classifier system for Arabic handwritten words recognition", *Pattern Recognition Letters*, vol. 93, pp. 123-132, 2017.
- [۲۹] J. Kennedy and R. Eberhart, "Swarm Intelligence", Morgan Kaufmann Publishers, 2001.
- [۳۰] S. Mirjalili, S. Z. M. Hashim and H. M. Sardroudi, "Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm", *Applied Mathematics and Computation*, Vol.218, No.22, pp. 11125-11137, 2012.
- [۳۱] E. Rashedi, H. Nazemabadi-poor, and S. Saryazdi, "GSA: A Gravitational Search Algorithm", *Information Sciences*, Vol.179, pp. 2232-2248, 2009.
- [۳۲] A. Boukharouba and A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits", *Applied Computing and Informatics*, Vol.13, No.1, Pages 19-26, 2017.
- [۳۳] R. C. Gonzalez, and R. E. Woods, "Digital image processing", Addison-Wesley, 2nd edition, 2002.