

بررسی شخصیت و خصوصیات فردی بر اساس دستنوشته فارسی با استفاده از درخت تصمیم

وحید قدس^۱ و حوریه عربیان^۲

چکیده

گرافولوژی علم یا فنی است که با مطالعه خط یک فرد، شخصیت و یا خلق و خوی و یا چگونگی روان او را می‌توان تشخیص داد. در این تحقیق، با اخذ دستخط و پرسشنامه شخصیتی MMPI از افراد مختلف یک پایگاه داده تشکیل شده است. ضمن استخراج ویژگی‌ها از دستخط، از درخت تصمیم به عنوان ابزار یادگیری ماشین جهت گروه بندی ویژگی‌ها استفاده شده است. سیستم پیاده سازی شده با دریافت نمونه دستخط روبش شده می‌تواند خصوصیات فردی نویسنده را پیش‌بینی نماید. نمونه‌های مورد بررسی شامل ۱۴۰ آزمودنی در دو گروه مختلف است. این دو گروه عبارتند از ۷۰ نفر از مددجویان زندان شهر سمنان و ۷۰ نفر از دانشجویان مرکز علمی کاربردی جهاد دانشگاهی سمنان. در راستای اخذ دستخط، یک نوشته‌ی واحد که متن آن از لحاظ تحریک احساسات خنثی و ضمنا ساده و روان باشد انتخاب شد و از آزمودنی‌ها خواسته شد که از روی آن در یک کاغذ A4 بی خط بنویسند. برای تشخیص نهایی از درخت تصمیم استفاده شده است. برای راستی‌آزمایی نتایج درخت تصمیم، از خروجی‌های پرسشنامه MMPI کمک گرفته شد که نتایج نویدبخشی به دست آمد. نتایج تشخیص‌درست ویژگی‌های استخراج شده توسط الگوریتم پیشنهادی از ۸۷٪ تا ۹۴٪ مطابقت با مشاهدات را نشان می‌دهد. نتایج تشخیص شخصیت حاصل از درخت تصمیم دقت ۶۹٪ را در مقایسه با نتایج پرسشنامه شخصیتی MMPI حاصل داد.

کلید واژه‌ها

گرافولوژی، دستخط، شخصیت، فارسی، درخت تصمیم.

۱ مقدمه

میلادی یک استاد فلسفه به نام "بالدی" و دو سال بعد کشیشی به نام "میشون" درباره شناسایی خلق و خوی و خصوصیات اخلاقی افراد، ضمن مطالعه خط و نوشته‌هایشان مطالبی اظهار داشتند و ادعا کردند که از خط یک شخص نیز می‌توان به چگونگی رفتار و شخصیت او پی برد و با خلق و خوی او آشنا شد.

یک خط شناس مانند صیادی است که به شکار علایم و نشانه‌ها می‌رود. هیچ دو دستخطی مثل اثر انگشت، دقیقا یکسان نیستند. دستخط هر کس، یگانه و منحصر به فرد است. الگوهای نوشتاری هیچکس با کس دیگر به هیچ وجه مطابقت ندارد.

خلق و خوی یک فرد در "من" که وی به هنگام صحبت به کار می‌رود خلاصه می‌شود و آن ثمره سخت پیچیده‌ای است که از میراث‌ها و عوامل فیزیولوژی، قبل و بعد از تولد و همچنین تجربه‌ها، دانش آموزشی‌ها و دانش اندوزی‌ها در

خط شناسی علمی را "گرافولوژی" می‌نامند و منظور علم و یا فنی است که با مطالعه خط یک فرد، شخصیت و یا خلق و خوی و یا چگونگی روان او را می‌توان تشخیص داد. در سال ۱۶۶۰

این مقاله در آبان‌ماه ۱۳۹۴ دریافت، در دیماه بازنگری و در بهمن ماه همان سال پذیرفته شد.

این مقاله طی قرارداد پژوهشی شماره ۱۷۴۷۹ مورخ ۹۲/۱۱/۲۷ بوسیله معاونت پژوهش و فناوری دانشگاه آزاد اسلامی واحد سمنان حمایت شده است که نویسندگان تشکر خالصانه خود را تقدیم می‌نمایند.

^{۲۰۱} گروه مهندسی برق، واحد سمنان، دانشگاه آزاد اسلامی

رایانامه: v.ghods@semnaniau.ac.ir

^۲ عضو مرکز علمی کاربردی جهاد دانشگاهی سمنان

رایانامه: tz_arabian@yahoo.com

نویسنده مسئول: وحید قدس

۲ پرسشنامه شخصیتی MMPI

پرسشنامه چند وجهی مینه سوتا (MMPI) در روان سنجی و مطالعات روان شناختی از جایگاه ویژه‌ای برخوردار است. از زمان تدوین پرسشنامه MMPI توسط هتاوی و مک کین لی در سال ۱۹۴۰ این آزمون یکی از پر مصرف ترین پرسشنامه‌های شخصیتی بالینی بوده است [۱۰]. اخوت و همکاران، ۷۱ عبارت از پرسشنامه MMPI را اقتباس و با در نظر گرفتن فرهنگ خاص ایرانی آن را در فرم کوتاهی تدوین و رشته ای از فعالیت های پژوهشی را در این زمینه آغاز کردند. پژوهش‌هایی که تاکنون انجام شده است، بیانگر ارزش تشخیصی فرم کوتاه ۷۱ جمله است [۱۱].

نمره‌گذاری این پرسشنامه در دو دسته از مقیاسها (مقیاسهای اعتباری و مقیاسهای بالینی) صورت می‌گیرد. مقیاسهای اعتباری (روایی) قبل از تفسیر و نمره‌گذاری مقیاسهای بالینی نمره گذاری و تفسیر می‌شود.

مقیاس‌های پرسشنامه MMPI به دو دسته تقسیم می‌شوند:

الف) مقیاس‌های اعتباری:

مقیاس L: این مقیاس شامل ۵ عبارت است که بیانگر خشکی، تحجر و در عین حال کمال‌گرایی آزمودنی است. نمره زیاد در این مقیاس بیان‌کننده آن است که آزمودنی سعی داشته چهره‌ای جز آنچه هست از خود معرفی کند.

مقیاس F: این مقیاس دارای ۱۵ عبارت است. نمره زیاد در این مقیاس بیشتر معلول اشتباه آزمودنی در پاسخ دادن به ۱۵ سوال آن، عدم فهم سوال‌ها و یا عدم همکاری وی است. افراد گوشه‌گیر و آنان که افکار غیر عادی دارند در این مقیاس نمره بالایی دارند. نمره بالا در این مقیاس نیز نشانه اغتشاش فکری آزمودنی و یا عدم همکاری عدم درک عبارات توسط آزمودنی است، در حالی که افراد آرام و قابل اعتماد در این مقیاس نمره پایینی به دست خواهند آورد.

مقیاس K: این مقیاس شامل ۱۶ عبارت است و بر اساس مطالعات تجربی تهیه شده است. در این مقیاس افرادی که خود را حقیر نشان می‌دهند و نیز افراد بیمار به سبب انکار علائم مرضی نمره بالایی کسب می‌کنند. به طور کلی این میزان برای اندازه‌گیری حالت دفاعی تهیه شده است. افراد بدبین و محتاط در این مقیاس نمره پایینی به دست خواهند آورد. با بررسی نمره آزمودنی در مقیاس‌های F و K می‌توان به نگرش‌های وی و چگونگی پاسخ دادن او به سوال‌هایی برد. هرگاه F در نیمرخ آزمودنی بالاتر از K باشد، وی سعی داشته است خود را بدتر از آنچه هست جلوه دهد. هرگاه F در نیمرخ آزمودنی پایین‌تر از k باشد، آزمودنی سعی داشته است خود را بهتر از آنچه که هست نشان دهد و درگیری‌های عاطفی خویش را انکار کند.

ب) مقیاس‌های بالینی:

مقیاس HS: این مقیاسکه متشکل از ۱۴ عبارت است مربوط به طرز تلقی و برداشت آزمودنی از وضع جسمانی خویش است.

او به وجود آمده است. یک خط شناس برای راه‌یابی و پیشبرد کار خود باید به خواص اصلی خلق و خوی و قوانینی که آنها را بهم مربوط می‌سازد، علم داشته باشد. بررسی یا امتحان یک نوشته (یک خط) علاوه بر جنبه علمی و تحقیقی آن، در بیشتر موارد، انگیزه‌ای دارد. آگاهی از این علم می‌تواند موارد استفاده‌ی زیاد و گوناگونی داشته باشد [۱]. گزینش متقاضیان استخدام، انتخاب همسر، کمک در تشخیص بیماری‌های روانی، از جمله کاربردهای گرافولوژی می‌باشد.

سیستم بیومتریک شناسایی نویسندگان خاص بر اساس گرافولوژی در [۲] ارائه شده است. دقت شناسایی به دست آمده با پنج خصیصه بر روی ۲۹ نویسنده، ۹۹٫۳۴٪ گزارش شده است. در [۳]، شبکه عصبی مصنوعی به عنوان طبقه بند مورد استفاده قرار گرفت. سپس، ادغام تصمیم برای شناسایی نویسنده مورد بهره برداری قرار گرفته است. در [۴]، برخی از ویژگی‌های جدید گرافولوژی معرفی شدند. ایده این بود که بخش‌های مهمی از امضا با استفاده از منحنی‌های Bezier شبیه‌سازی شوند و سپس استخراج ویژگی‌ها از آن‌ها صورت پذیرفته است.

از آنجا که گرافولوژی بر مبنای ویژگی‌های دستنوشته است، زبان دستنوشته در استخراج ویژگی‌ها مؤثر است. بنابراین، برای هر زبان از جمله زبان فارسی استخراج ویژگی‌های مناسب دیداری برای گرافولوژی با استفاده از کامپیوتر امری ضروری است. در زمینه گرافولوژی دستنوشته فارسی کارهای اندکی انجام شده است. برای اولین بار گرافولوژی دستنوشته فارسی به کمک کامپیوتر توسط احسان الله کبیر و علی بهرامی شریف بررسی شد [۵]. پس از آن افراد دیگری روی استخراج ویژگی‌های دستخط فارسی کار کردند [۶] و برخی از محققین ضمن استخراج ویژگی از دستخط با استفاده از طبقه بندها مانند شبکه‌های عصبی مصنوعی و سیستم استنتاج فازی نسبت به دسته بندی ویژگی‌ها و تشخیص خصوصیات روانشناختی از روی دستخط اقدام نمودند [۷ و ۸ و ۹].

در بخش دوم مقاله راجع به پرسشنامه شخصیتی MMPI توضیح داده می‌شود. در بخش سوم، پایگاه داده ارائه شده است. در بخش چهارم، نحوه تصمیم‌گیری با استفاده از پرسشنامه MMPI شرح داده می‌شود. استخراج ویژگی‌های هر دستخط و بررسی آن‌ها در بخش پنجم بحث می‌شود. در بخش ششم، گروه بندی با استفاده از درخت تصمیم آورده شده است. نتایج و مقایسه با کارهای مشابه در بخش هفتم و نتیجه‌گیری و پیشنهادات در بخش هشتم بررسی و ارائه شده است.

۳ پایگاه داده

هدف از انجام این تحقیق ارزیابی سیستم هوشمند گرافولوژی دستنوشته فارسی می‌باشد. افراد گروه نمونه شامل دو گروه مختلف می‌باشد این دو گروه عبارتند از [۱۳]:

یک گروه عادی از میان دانشجویان مقطع کاردانی و کارشناسی مرکز آموزش عالی علمی کاربردی جهاد دانشگاهی سمنان که بصورت تصادفی انتخاب شدند. تعداد افراد این گروه ۷۰ نفر شامل ۲۵ مرد و ۴۵ زن که اغلب بین ۲۰ تا ۳۰ سال سن داشتند، می‌باشند.

گروهی دیگر، مددجویان زندان شهر سمنان بودند. تعداد افراد این گروه ۷۰ نفر مرد در بازه سنی ۲۱ تا ۵۷ سال می‌باشند. اخذ پرسشنامه و دستخط از این گروه با همکاری روانشناس زندان صورت گرفته است تا حتی المقدور از لحاظ اعتباری قابل اعتماد باشد.

از آنجائیکه رعایت مسایل اخلاقی و تضمین حقوق انسانی آزمودنی‌ها یکی از عمده ترین مسایل در پژوهش‌های آموزشی و پرورشی است. اولین و مهمترین حقی که هر انسان دارد عبارت است از آزادی عدم مشارکت در پژوهش. این اصل یا آزادی از این ناشی می‌شود که انسان علاقمند است برخی از اطلاعات مربوط به خود را از اعضای جامعه دور نگاه دارد یا این گونه اطلاعات را به هیچ وجه در اختیار آنها قرار ندهد [۱۴].

لذا طی فرایند اخذ پرسشنامه و دستخط افراد در صورت تمایل در این تحقیق همکاری نمودند به عبارتی در مشارکت یا عدم مشارکت آزاد بودند. ضمناً تنها کسانی که تمایل به دانستن نتیجه پرسشنامه بودند اسامی خود را ذکر کردند و برای سایر افراد تنها از کد استفاده گردید. آزمودنی‌ها ضمن نوشتن یک متن واحد روی کاغذ بدون خط پرسشنامه شخصیتی ۷۰ سوالی MMPI را تکمیل نمودند.

دستخط نمایش داده شده در شکل ۱، یک دستخط ریز است و تنها از نیمی از کاغذی که در اختیار داشته است استفاده نموده است. نوشته‌ی خیلی ریز نشان دهنده‌ی نیروی زندگی ضعیف، فقدان جنب و جوش، خست و کم‌رویی می‌باشد [۱]. با نگاهی به نمودار T مربوطه که در شکل ۲ نمایش داده شده است، برافراشتگی بالا در مقیاسهای PT و HS دیده می‌شود. برافراشتگی بالا در مقیاس PT در نمودار T نشان دهنده خستگی روانی، ضعف روحی، اعتماد به نفس پایین و حساسیت بی‌مورد و برافراشتگی بالا در مقیاس HS در نمودار فوق نشان دهنده خود بیمار انگاری و اضطراب و اندیشه بیهوده در مورد سلامتی می‌باشد.

افرادی که از وضع جسمانی خود شکایت دارند در این مقیاس نمره بالایی خواهند آورد.

مقیاس D: این مقیاس که شامل ۲۰ عبارت است حالات ناامیدی، گوشه گیری و همچنین عدم اعتماد نفس افراد را می‌سنجد.

مقیاس HY: این مقیاس متشکل از ۲۵ عبارت است. افرادی که زیاد متکی به دیگران هستند و در مورد ناراحتی‌های خود به اندازه کافی بینش ندارند و کسانی که رفتارشان فاقد پذیرش اجتماعی است در این میزان نمره‌های بالایی به دست می‌آورند.

مقیاس PD: این مقیاس دارای ۱۹ عبارت است و نشان دهنده عدم توافق و ارتباط گرم با اجتماع و خانواده است. نمره زیاد در این میزان مربوط به افرادی است که بر علیه خانواده و محیط عصیان می‌کنند. نوجوانان بزهکار نیز در این مقیاس نمره بالایی به دست می‌آورند.

مقیاس PA: این مقیاس دارای ۱۴ عبارت است که محتوای آن‌ها را حساسیت‌ها، موازین اخلاقی، بدبینی و شکایت از دیگران و اشاره به کمبودهای آنها تشکیل می‌دهد. این مقیاس افکار پارانوید را باز می‌شناسد. نمره بالا در این مقیاس علاوه بر سوء ظن، بیانگر این موضوع است که فرد گرایش‌های خود آزاری خویش را بیرونی کرده و متوجه دیگران می‌سازد.

مقیاس PT: این مقیاس دارای ۱۶ عبارت است که محتوای آنها را علائم مرضی اضطراب و تشویش، ترس‌ها و دلهره‌های بیجا و بی‌ارزش فرد تشکیل می‌دهد. افرادی که وسواسی بوده و از پدیده‌ها وحشت دارند و در اتخاذ تصمیم، ناتوانایی‌هایی از خود نشان می‌دهند و قادر نیستند فکر خود را بر روی موضوعی متمرکز کنند در این مقیاس نمره بالایی به دست می‌آورند. بین این مقیاس و مقیاس SC همبستگی بالایی وجود دارد. اگر فردی در این دو مقیاس نمره بالایی بگیرد وجود اسکیزوفرنیا را در مراحل پیشرفته نشانی می‌دهد.

مقیاس SC: این مقیاس شامل ۲۰ عبارت است که محتوای آنها را گوشه‌گیری، افکار عجیب و غیر عادی، شکایت از وضع خانوادگی و چگونگی برداشت از خود تشکیل می‌دهد. فردی که خود را غریبه، یکه و تنها احساس می‌کند و معتقد است که عضو جامعه خود نیست و مردم او را درک نمی‌کنند در این مقیاس نمره بالایی می‌آورد. البته بعضی از افراد نورتیک نیز در این مقیاس نمره بالایی می‌آورند.

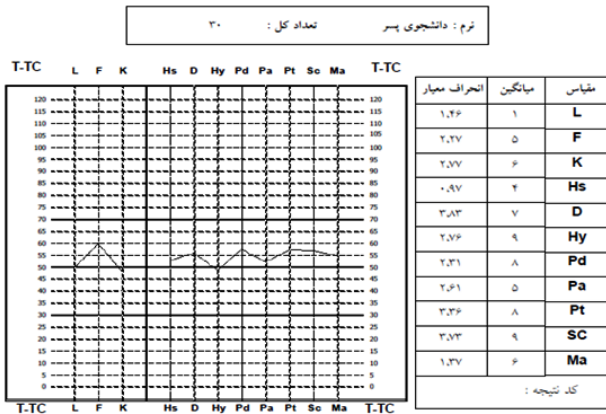
مقیاس MA: این مقیاس دارای ۱۱ عبارت است و دامنه وسیعی از فعالیت‌هایی که متوجه امور بیهوده و بر خلاف سنن اجتماعی هستند نشان می‌دهند. افرادی که رفتارهای خود را در ضبط و کنترل ندارند معمولاً در این مقیاس نمره بالایی به دست می‌آورند. این افراد خوش بینی غیر منطقی نسبت به آینده خود داشته و بسیار حساس و زود رنج هستند [۱۲].

۴ نتایج کلی پرسشنامه MMPI

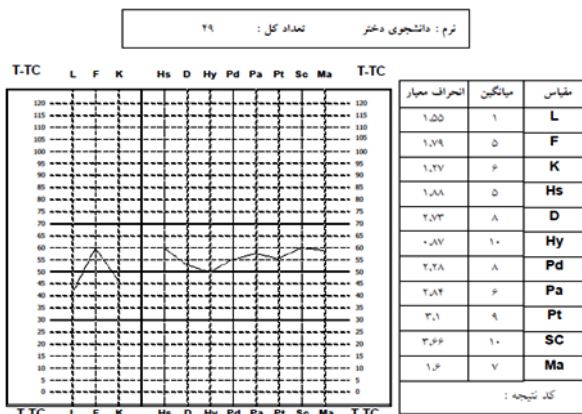
در این بخش به نتایج کلی پرسشنامه های اخذ شده از مددجویان زندان شهر سمنان و دانشجویان به تفکیک مرد و زن نگاهی می-شود.

شکل ۷ نمودار T نتیجه کلی مددجویان زندان شهر سمنان را نشان می-دهد. آنچه نمودار نشان می-دهد، برافراستگی بالا در مقیاس های D، Hy، Pd، Pa، Sc و Ma و برافراستگی پایین در مقیاس Ma می-باشد. برافراستگی پایین در Ma نشان دهنده پایین بودن خودشیفتگی، بالا بودن مقیاس D نشان دهنده بالا بودن افسردگی و بد بینی می-باشد. بالا بودن مقیاس Hy مربوط به کسانی است که رفتارشان فاقد پذیرش اجتماعی می-باشد این مقیاس برای مددجویان زندان دارای برافراستگی بالاست در صورتی که برای دانشجویان (شکلهای ۸ و ۹) این مقیاس دارای برافراستگی پایین می-باشد.

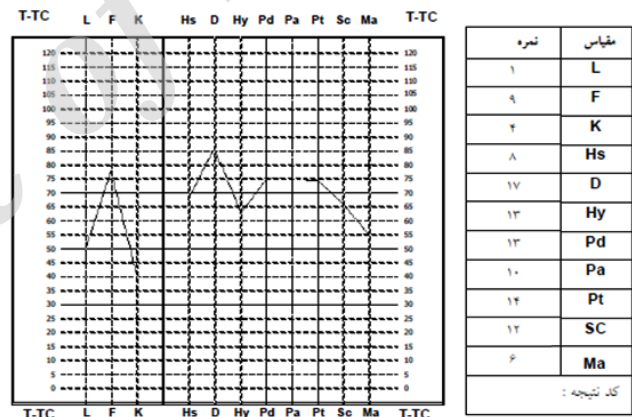
بالا بودن مقیاس Pd نشان دهنده بی تفاوتی و ناسازگاری اجتماعی می-باشد و بالا بودن مقیاس Pa نشان دهنده عدم اعتماد و بالا بودن مقیاس Sc نشان دهنده اختلال هویت خویش و انزواجویی می-باشد.



شکل ۸ نتیجه کلی پرسشنامه دانشجویان پسر



شکل ۹ نتیجه کلی پرسشنامه دانشجویان دختر



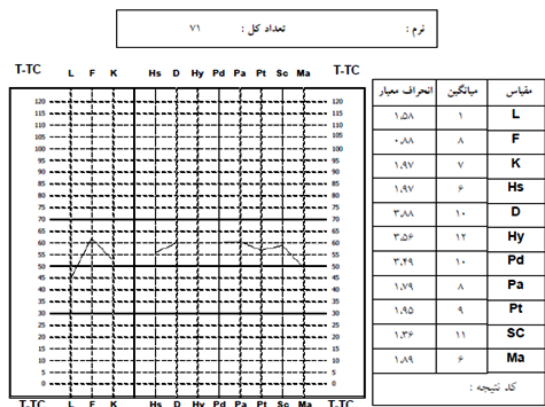
شکل ۶ نتیجه پرسشنامه (نمودار T) مربوط به شکل ۵

۵ استخراج ویژگی ها از دستخط

ورودی سیستم تصویر دستخط های اخذ شده از نمونه جامعه آماری می-باشد که با درجه تفکیک ۶۰۰ dpi رویش شده است و بعد از نرمال سازی ۱۰۰ dpi می-شود.

۵-۱ پیش پردازش

در مرحله پیش پردازش ابتدا تصویر خوانده شده نرمال سازی می-گردد تا برنامه نوشته شده برای کلیه تصاویر با اندازه های مختلف قابل استفاده باشد. تمام صفحات مجموعه داده این تحقیق صفحه A4 می-باشد. برای یکسان سازی، تصویر خوانده شده به اندازه ۱۱۸۰*۸۲۸ تغییر اندازه (dpi ۱۰۰) داده می-شود تا اگر اندازه کاغذی که رویش شده و یا دقت روبشگر مانند نمونه های آموزشی نباشد، تاثیری در نتیجه حاصل به وجود نیاید. سپس تصویر دودویی می-شود. دودویی کردن، سطوح روشنایی تصویر را به دو سطح تبدیل می-کند. بدین طریق، انجام سایر عملیات به سادگی روی آنها انجام خواهد شد. در بسیاری از موارد تصویر نمونه ای که رویش شده است دارای عوارض و لکه های ناخواسته-ای هستند که نویز نامیده می-شوند. این نویزها در هنگام رویش تصویر، کثیفی کاغذ و یا نوک سر قلم به وجود می-آید. برای حذف این نویزها از فیلتر پایین گذر یا میانگین استفاده شده است. فیلتر



شکل ۷ نتیجه کلی پرسشنامه مددجویان شهر سمنان

۵-۵ شیب خط

تصویر ورودی با زوایای مختلف به فاصله ۰,۵ درجه، در بازه ۴۵- درجه تا ۴۵ درجه چرخانده می‌شود. پس از هر چرخش، افکنش افقی آن محاسبه می‌شود. سپس، آنتروپی این افکنش به دست می‌آید. زاویه ای که این آنتروپی را به کمترین مقدار خود می‌رساند، بیانگر شیب خط است [۵].

۶-۵ بررسی نمونه‌هایی از دستنوشته‌ها و خروجی الگوریتم استخراج ویژگی

در این بخش، خروجی الگوریتم استخراج ویژگی برای نمونه‌هایی از دستخط و علت خطای الگوریتم در این نمونه‌ها بررسی شده است. صحت خروجی الگوریتم بر اساس داده‌های آموزشی تعریف شده در جدول ۱ می‌باشد.

در جدول ۱، با توجه به تجربه و مشاهدات، فرضیات ذیل در نظر گرفته شده است. چنانچه عدد مربوط به کجی متن (زاویه خطوط با خط افقی) در خروجی الگوریتم بین ۱۶ تا ۶۰ باشد، خطوط دستنوشته پایین رونده در نظر گرفته شده است. چنانچه این عدد بین ۰ تا ۱۵ باشد، دستنوشته صاف در نظر گرفته شده است. اگر عدد مربوط به حاشیه‌ها در خروجی الگوریتم برابر یا بیشتر از ۸۷ باشد، متن با حاشیه زیاد در نظر گرفته شده است. اگر این عدد کمتر از ۷۲ باشد، حاشیه متن کم و چنانچه بین ۷۳ تا ۸۶ باشد حاشیه متن معمولی می‌باشد. اگر عدد مربوط به درشتی کلمات در خروجی الگوریتم برابر یا بیشتر از ۳۲۵ باشد، دستخط درشت در نظر گرفته شده است. اگر این عدد زیر ۱۹۰ باشد، خط ریز و چنانچه این عدد بین ۱۹۱ تا ۳۲۴ باشد، دستخط متوسط می‌باشد.

جدول ۱ داده‌های آموزشی

دسته	کجی متن	حاشیه ها	درشتی کلمات
۱	۱۶ تا ۶۰	۷۲ تا	از ۳۲۵ به بالا
۲	۱۶ تا ۶۰	از ۸۷ به بالا	از ۳۲۵ به بالا
۳	۱۶ تا ۶۰	۷۲ تا	تا ۱۹۰
۴	۱۶ تا ۶۰	از ۸۷ به بالا	تا ۱۹۰
۵	۰ تا ۱۵	۷۲ تا	از ۳۲۵ به بالا
۶	۰ تا ۱۵	از ۸۷ به بالا	از ۳۲۵ به بالا
۷	۰ تا ۱۵	۷۲ تا	تا ۱۹۰
۸	۰ تا ۱۵	از ۸۷ به بالا	تا ۱۹۰
۹	۱۶ تا ۶۰	۷۲ تا	۱۹۱ تا ۳۲۴
۱۰	۱۶ تا ۶۰	از ۸۷ به بالا	۱۹۱ تا ۳۲۴
۱۱	۱۶ تا ۶۰	۷۳ تا ۸۶	از ۳۲۵ به بالا
۱۲	۱۶ تا ۶۰	۷۳ تا ۸۶	تا ۱۹۰
۱۳	۰ تا ۱۵	۷۲ تا	۱۹۱ تا ۳۲۴
۱۴	۰ تا ۱۵	از ۸۷ به بالا	۱۹۱ تا ۳۲۴
۱۵	۰ تا ۱۵	۷۳ تا ۸۶	از ۳۲۵ به بالا
۱۶	۰ تا ۱۵	۷۳ تا ۸۶	تا ۱۹۰

پایین گذر دو تاثیر روی تصویر می‌گذارد: حذف نویز انجام می‌دهد و همچنین روی تصویر محو شدگی اتفاق می‌افتد. به این ترتیب لبه‌ها محو شده و تا حدی گسستگی‌ها از بین می‌رود. سپس از عملیات ریخت‌شناسی جهت حذف نویز بهتر و از بین رفتن گسستگی‌ها استفاده شده است. پیوسته کردن خطوط به کاهش خطا در استخراج ویژگی خصوصاً محاسبه تعداد خط کمک شایانی می‌نماید [۶].

۲-۵ تعداد خطوط اصلی نوشتار

برای یافتن تعداد خطوط دستنوشته از الگوریتم مرجع [۶] استفاده شده است. ابتدا، "عرض قلم" را به دست می‌آوریم. به طوری که در افکنش افقی و عمودی، تعداد نقاط سیاه تمام اجزای پیوسته شمارش می‌شود. سپس، عدد دارای فراوانی بیشتر به عنوان عرض قلم تعیین می‌گردد. در ادامه الگوریتم، با جریانسازی، ساختار کلی دستنوشته مشخص می‌شود. برای این منظور، ابتدا نواحی کم مساحت که معمولاً سرکش‌ها و علامت مد و غیره هستند، و ممکن است در تمایز خطوط مشکل ایجاد کنند، حذف می‌شوند. سپس از عنصر ساختاری مستطیل با ابعاد ۵ و ۷ برابر عرض قلم برای گسترش کلمات استفاده شده است تا هر کلمه به یک قطعه پیوسته تبدیل شود. سپس با اعمال جریانسازی بر روی این قطعات، سعی شده تا با متصل کردن قطعات مرحله قبل به یکدیگر، ساختار کلی هر خط از دستنوشته ایجاد شود. پس از آن، هر خط از دستنوشته به صورت یک قطعه پیوسته خواهد بود. با برچسب‌گذاری، خط‌ها از یکدیگر متمایز شده و قابل شمارش هستند.

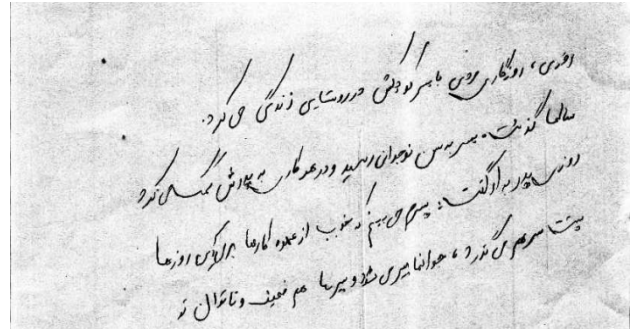
۳-۵ حاشیه متن

پس از به دست آمدن هر خط از دستنوشته به صورت یک قطعه پیوسته، حداقل شماره ستون پیکسل آغازین هر خط را ذخیره کرده و به عنوان ویژگی حاشیه در نظر گرفته می‌شود. از آنجایی که در گرافولوژی علاوه بر میزان حاشیه، از تفاوت بین حاشیه چپ و حاشیه راست نیز برداشت‌هایی در مورد خلق و خوی افراد می‌توان داشت، حاشیه چپ و راست جداگانه محاسبه و هر کدام به عنوان یک ویژگی جداگانه و میانگین حاشیه چپ و راست نیز به عنوان یک ویژگی دیگر مورد استفاده قرار گرفت.

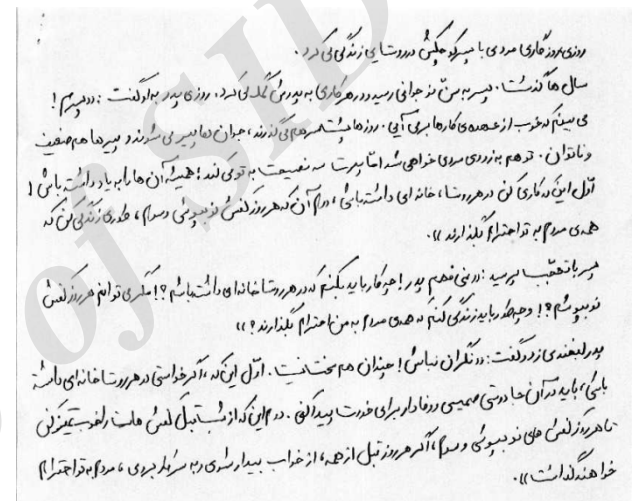
۴-۵ درشتی خط

برای تشخیص میزان درشت نویسی کلمات، ابتدا بخش‌های پیوسته حاصل از گسترش ریخت‌شناسی برچسب زده می‌شوند و مساحت هر بخش محاسبه می‌شود. سپس بخش‌های با مساحت کمتر از ۲۰ برابر مربع عرض قلم حذف می‌شوند. میانگین مساحت‌های باقیمانده تقسیم بر عرض قلم، به عنوان معیار برای درشت‌نویسی لحاظ می‌شود.

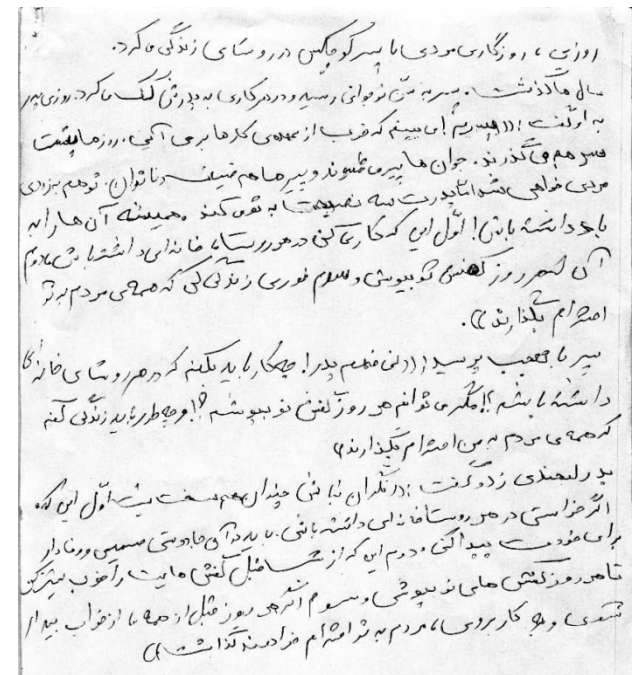
شکل ۱۰ دستنوشته‌ای با ۴ خط متن، جهت پایین رونده، حاشیه زیاد و نسبتاً درشت می‌باشد. در جدول ۲ می‌بینیم اعداد مربوط به استخراج ویژگی‌ها مناسب می‌باشند. شکل ۱۱ نوشته‌ای با ۱۲ خط متن، کمی متن زاویه‌دار، درشتی کلمات معمولی و با حاشیه‌ای کم می‌باشد.



شکل ۱۰ نمونه‌ای از دستخط با جهت پایین و حاشیه زیاد



شکل ۱۱ نمونه‌ای از دستخط با جهت پایین و حاشیه کم



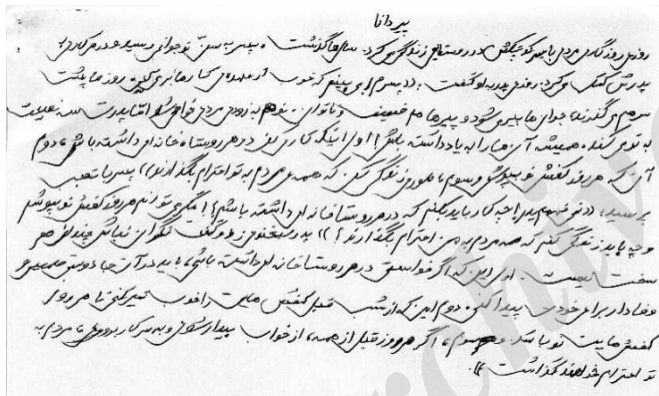
شکل ۱۲ نمونه‌ای از دستخط درشت و حاشیه کم

در جدول ۲ می‌بینیم اعداد مربوط به تعداد سطر، حاشیه متن، درشتی کلمات و زاویه متن بسیار مناسب می‌باشند. شکل ۱۲ نوشته‌ای با ۱۶ خط متن، درشت با حاشیه‌ای کم می‌باشد. در جدول ۲ دیده می‌شود که اعداد مربوط به حاشیه متن، درشتی کلمات و زاویه متن مناسب می‌باشند.

جدول ۲ خروجی الگوریتم استخراج ویژگی برای نمونه‌ای از دستنوشته‌ها

شماره شکل دستخط	تعداد سطر	حاشیه متن	درشتی کلمات	زاویه متن
۱۰	۴	۱۲۰	۲۸۱	۲۴,۵
۱۱	۱۲	۶۱	۲۰۵	۹,۵
۱۲	۱۶	۶۴	۳۲۹	۱,۰
۱۳	۱۱	۳۸	۱۵۴	۲,۵

شکل ۱۳ نوشته‌ای با ۱۲ خط متن، ریز با حاشیه‌ای کم می‌باشد. در جدول ۲ مشاهده می‌شود که اعداد مربوط به تعداد سطر، حاشیه متن، درشتی کلمات و زاویه متن مناسب می‌باشند. تنها خطای الگوریتم در محاسبه تعداد خط است. تنها خطای الگوریتم در تعداد خط می‌باشد که علت آن با توجه به الگوریتم تعداد خط، چسبیدن خطوط ۷ و ۸ می‌باشد.



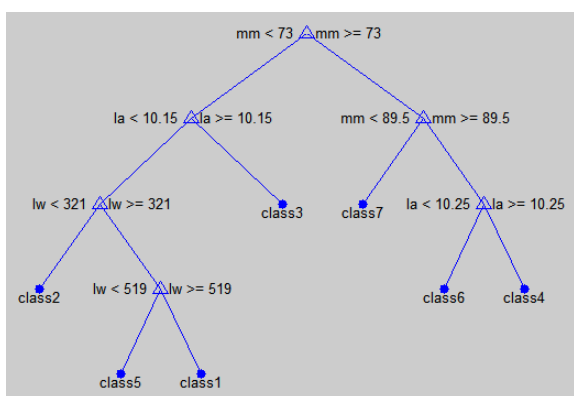
شکل ۱۳ نمونه‌ای از دستخط ریز و حاشیه کم

۶ گروه بندی ویژگی‌ها با درخت تصمیم

یادگیری درخت تصمیم گیری یکی از پرکاربردترین و کارآمدترین روشهای یادگیری استقرایی است. این روش در یادگیری توابع گسسته مقدار با داده های خطا دار به کار می‌رود [۱۵]. استقرا ایجاد قوانین کلی از مجموعه مثالها است. مثالها در قالب جداولی نمایش داده می‌شود که به آنها جداول تصمیم گیری گفته می‌شود. از طریق مثالها دانش استخراج می‌شود و در قالب درختی به نام درخت تصمیم گیری بازنمایی خواهند شد [۱۶].

در درخت تصمیم هر گره داخلی، ویژگی‌ای از نمونه را آزمایش می‌کند و هر شاخه‌ای که از آن گره خارج می‌شود متناظر یک مقدار ممکن برای آن ویژگی می‌باشد. همچنین هر گره برگ، بیانگر یک دسته بندی است. هر نمونه با شروع از گره ریشه درخت و آزمایش ویژگی مشخص شده

به ترتیب ردیف‌ها به دسته مربوط اختصاص دادیم و نهایتاً از دستور classregtree برای طبقه بندی داده های تست استفاده شده است. ۷۰٪ نمونه‌ها برای آموزش و ۳۰٪ نمونه‌ها برای آزمون در نظر گرفته شده است و یکبار هم نمونه‌های آموزش و آزمون (با توجه به تعداد) جایجا شده است و میانگین نتایج محاسبه شد. پس از آموزش الگوریتم، هربردار دلخواه بر اساس این درخت طبقه بندی می‌گردد. شکل ۱۴ درخت تصمیم حاصل از شبیه سازی را نشان می‌دهد.



شکل ۱۴ درخت تصمیم حاصل از شبیه سازی

۷ نتایج

بر اساس دستنوشته های جمع آوری شده، ویژگی‌هایی که بر اساس الگوریتم پیشنهادی استخراج شده‌اند و پرسشنامه شخصیتی MMPI، چندین آزمایش انجام شد که در این بخش به شرح آن‌ها می‌پردازیم.

۷-۱ تشخیص فرد خبره بر روی ویژگی‌ها

از مقایسه نتایج تشخیص ویژگی‌های استخراج شده از الگوریتم پیشنهادی و فرد خبره بر روی ۱۴۳ نمونه دستنوشته، جدول ۳ به دست آمد. این جدول نشان می‌دهد که در مرحله استخراج ویژگی نیز اشکالاتی در به دست آوردن ویژگی‌ها رخ می‌دهد.

جدول ۳ بررسی نتایج استخراج ویژگی

معیار استفاده شده در گرافولوژی	استخراج درست
کجی متن	۸۷٪
تعداد خط	۹۴٪
حاشیه	۹۰٪
درستی خط	۸۸٪

۷-۲ تشخیص انسانی

در بررسی دیگری، از بیست نفر از افرادی که نسبت به شخصیت آن‌ها شناخت وجود داشت، خواسته شد که متن مورد نظر را روی یک کاغذ بی‌خط بنویسند. سپس، خروجی درخت تصمیم با

توسط این‌گروه و حرکت در شاخه متناظر با مقدار ویژگی داده شده در نمونه، دسته بندی می‌شود. این فرآیند برای هر زیر درختی که گره جدید، ریشه آن می‌باشد تکرار می‌شود. درخت تصمیم نمونه‌ها را به نحوی دسته بندی می‌کند که از ریشه به سمت پایین رشد می‌کنند و در نهایت به گره‌های برگ می‌رسد [۱۷].

بر اساس دریافتی که از مرجع [۱] حاصل شد، با توجه به ویژگی‌های استخراج شده شامل درستی خط، حاشیه و زاویه متن، می‌توان خصوصیات متقابل زیر را پی برد:

* افسرده- توانایی کنترل احساسات

* برون‌گرا- درون‌گرا- نرمال از لحاظ درون‌گرایی و برون‌گرایی

* فعال و متحرک- ضعیف و کم‌رو- نرمال از لحاظ جنب و جوش

که از خصوصیات فوق، ۱۸ دسته مختلف بوجود می‌آید. با توجه به اینکه در برخی از دسته‌ها نمونه دستخط کافی در مجموعه پایگاه داده وجود نداشت، حالت‌هایی که زیر ۱۰ نمونه دستخط بودند، با توجه به احتمال فراوانی کم آنها در جامعه عادی، حذف گردیدند و برای اجرای الگوریتم درخت تصمیم، از جدول ۲ استفاده شده است.

جدول ۲ گروه بندی شخصیت افراد مورد استفاده در شبیه سازی

شخصیت	دسته
احساساتش را می‌تواند کنترل کند، برون‌گرا، فعال و	۱
احساساتش را می‌تواند کنترل کند، برون‌گرا،	۲
افسرده، برون‌گرا، نرمال از لحاظ جنب و جوش	۳
افسرده، درون‌گرا، نرمال از لحاظ جنب و جوش	۴
احساساتش را می‌تواند کنترل کند، برون‌گرا، نرمال	۵
احساساتش را می‌تواند کنترل کند، درون‌گرا، نرمال	۶
احساساتش را می‌تواند کنترل کند، نرمال از جنبه	۷

در این تحقیق، جهت ساخت درخت تصمیم از تابع classregtree در نرم افزار متلب استفاده شده است. به ازای هر گروه شخصیتی ده نمونه آموزشی انتخاب شد. با توجه به اینکه شخصیت افراد را به هفت گروه دسته‌بندی کرده‌ایم، ۷۰ داده آموزشی را به ترتیب با استفاده از دستور meas در سه ستون و ۷۰ ردیف وارد کردیم (سه ستون مربوط به سه ویژگی استخراج شده می‌باشد). سپس، با استفاده از دستور species در داخل یک حلقه، دسته‌های ۱ تا ۷ را

[۶ و ۵]. البته این تحقیقات در خصوص استخراج ویژگی بسیار ارزشمند بوده‌اند، خصوصاً مقاله [۵] که پایه و اساس اغلب تحقیق‌های گرافولوژی در دستنوشته‌های فارسی می‌باشد. در تحقیق‌های [۷، ۸ و ۹] از ابزارهای یادگیری ماشین برای دسته بندی و تشخیص خصوصیات روانشناختی افراد استفاده نمودند. در تحقیق [۱۸] تنها تاثیر حالت روحی عجله بروی دستخط با استفاده از شبکه عصبی بررسی شده است و در تحقیق [۹]، تنها شیب خط برای تشخیص هیجانان و احساسات افراد با کمک سیستم استنتاج فازی بررسی شده است. در تحقیق [۸]، تشخیص خصوصیات افراد با کمک سیستم استنتاج فازی صورت گرفته است. در هیچ کدام از تحقیق‌های صورت گرفته فوق کاری در راستای راستی آزمایی صورت نگرفته است. در تحقیق [۷]، پیش بینی خصوصیات روانشناختی افراد از روی دستخط با کمک شبکه عصبی صورت گرفته است و از پرسشنامه شخصیتی MMPI جهت مدل کردن سیستم استفاده نموده است. ما در این تحقیق برای اولین بار از درخت تصمیم به منظور دسته‌بندی ویژگی‌های استخراج شده از دستخط استفاده نمودیم و از میان کارهای انجام شده مذکور، این تحقیق تنها تحقیقی است که از یک پرسشنامه شخصیتی به منظور راستی‌آزمایی الگوریتم خود استفاده نموده است.

۸ نتیجه‌گیری و پیشنهادات

در این تحقیق، یک پایگاه داده دستنوشته فارسی برای تشخیص شخصیت افراد معرفی گردید. استفاده از یک پرسشنامه به همراه اخذ دستنوشته به منظور راستی‌آزمایی از خصوصیات این مجموعه داده است. یک متن واحد و خنثی به آزمودنی‌ها به منظور عدم تاثیر متن بر جریحه دار کردن احساسات داده شد. داشتن دو دسته جامعه آماری امکان مقایسه کلی و به دست آمدن نتایج روانشناختی را میسر می‌سازد. سپس، استخراج ویژگی از دستنوشته‌ها صورت پذیرفت. از درخت تصمیم به منظور دسته بندی و تشخیص خصوصیات روانشناختی استفاده شد. نتایج حاصل از بازشناسی با خروجی‌های پرسشنامه MMPI مقایسه گردید که رضایتبخش می‌باشد.

برای کارهای آتی پیشنهاد می‌شود چنانچه امکان دسترسی به جامعه آماری باشد که بتوانند محقق را در تکمیل پرسشنامه‌های ۱۸۷ سوالی یا ۲۵۰ سوالی یاری نمایند، میتوان از پرسشنامه‌های شخصیتی کتل و نئو استفاده نماید. هرچند در این تحقیق از راهنمایی‌های اساتید روانشناسی و مشاوران استفاده کردیم، ولی نیاز به همکاری یک مرکز مشاوره به طور مستمر احساس می‌شود. به طوری که به جای جامعه آماری تصادفی، داشتن نمونه‌های کافی از طبقه بندی‌های شخصیتی را امکان پذیر نماید.

شناختی که از شخصیت آن‌ها وجود داشت، مقایسه شد. تشخیص شخصیت افراد مورد آزمایش و شناخت قبلی از شخصیت آنان، دقت ۷۸٪ برای الگوریتم پیاده سازی شده را تایید می‌نماید.

۳-۷ راستی‌آزمایی با پرسشنامه شخصیتی MMPI

در بررسی نهایی، نتایج حاصل از پیاده سازی استخراج ویژگی و درخت تصمیم با نتایج حاصل از پرسشنامه شخصیتی MMPI مقایسه شده است و این مقایسه دقت ۶۹٪ را برای پیاده سازی تایید می‌نماید.

با اینکه پرسشنامه MMPI بارها در نقاط مختلف جهان استفاده شده است و مورد اطمینان است [۱۰]، ممکن است یکی از دلایل کاهش دقت هنگام مقایسه با پرسشنامه شخصیتی MMPI، درست پاسخ ندادن سهوی یا عمدی برخی افراد به سوالات پرسشنامه و عدم تطبیق کامل بعضی مقیاس‌های پرسشنامه و طبقه بندی‌های شخصیتی این تحقیق می‌باشد. در ضمن، پرسشنامه MMPI بیشتر اختلال شخصیت را تشخیص می‌دهد. به عنوان مثال، ممکن است کسی که دستخط پایین رونده دارد، حالات افسردگی مانند غمگینی، بی‌حوصلگی و اهمال کاری داشته باشد ولی دچار اختلال افسردگی نباشد. لذا در نمودار T حاصل از نتیجه پرسشنامه شخصیتی MMPI، مقیاس D که نشان‌دهنده افسردگی می‌باشد برافراشتگی نداشته باشد.

بنا به دلایل زیر بهتر است در کارهای آتی اخذ نمونه دستخط با همکاری یک مرکز مشاوره صورت پذیرد:

(الف) شناخت شخصیت از مجموع تست، مشاهده و مصاحبه بالینی صورت می‌پذیرد و وجود یک پرسشنامه به تنهایی، نمی‌تواند شخصیت فرد را بطور دقیق تشخیص دهد.

(ب) اخذ بیش از ۱۰ نمونه دستخط از هر طبقه شخصیت. در غیر اینصورت با وجود داشتن تعداد زیادی نمونه دستخط، هنگام تقسیم آنها به طبقه‌های مختلف شخصیت، امکان عدم وجود مقدار کافی نمونه دستخط از هر طبقه شخصیتی وجود دارد.

۴-۷ مقایسه با دیگر روش‌ها

آنجا که گرافولوژی بر مبنای ویژگی‌های دستنوشته است، زبان دستنوشته در استخراج ویژگی‌ها مؤثر است. بنابراین، برای هر زبان از جمله زبان فارسی استخراج ویژگی‌های مناسب دیداری برای گرافولوژی با استفاده از کامپیوتر امری ضروری است [۵]. در زمینه گرافولوژی دستنوشته فارسی کارهای اندکی انجام شده است که در این قسمت به مقایسه این کارها با این تحقیق پرداخته شده است.

در برخی تحقیقاتی که در زمینه گرافولوژی صورت گرفته است، تنها به استخراج ویژگی از دستنوشته و نهایتاً مقایسه ویژگی‌های استخراج شده با نظر افراد خبره اکتفا نمودند ولی هیچگونه دسته بندی با استفاده از ابزارهای یادگیری ماشین انجام نشده است

تشکر و قدردانی

در پایان از جناب آقای دکتر سید موسی طباطبایی که در جمع آوری داده از زندانی ها، تنها با انگیزه کمک علمی، ما را یاری کردند و از مسئولین محترم مرکز آموزش عالی علمی کاربردی جهاد دانشگاهی سمنان که در توزیع پرسشنامه ها و اخذ دستخط از دانشجویان ما را یاری کردند، تشکر و قدردانی می نمایم. همچنین از معاونت پژوهش و فناوری دانشگاه آزاد اسلامی واحد سمنان نیز تشکر و قدردانی می گردد که این تحقیق از نتایج طرح پژوهشی و با حمایت این واحد انجام شده است.

مرجع

- [۱] بوشاتو، گ.؛ ترجمه و پژوهش احمد یلدا، خط و شخصیت، کتابسرا تهران، ۱۳۷۶.
- [2] O. Santana, C.M. Travieso, J.B. Alonso, M.A. Ferrer. Writer identification based on graphology techniques. IEEE Aerospace and Electronic Systems Magazine. 25 (2010) 35-42.
- [3] J.L. Vasquez, C.M. Travieso, J.B. Alonso. Off line writer identification based on graphometric parameters IEEE 17th International Conference on Intelligent, Engineering Systems (INES) USA, 2013. p. 219-23.
- [4] D. Bertolini, L.S. Oliveira, E. Justino, R. Sabourin. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. Pattern Recognition. 43 (2010) 387-96.
- [۵] بهرامی شریف، ع.؛ کبیر، ا.؛ "گرافولوژی دست نوشته فارسی به کمک کامپیوتر"، نشریه مهندسی برق و مهندسی کامپیوتر ایران، پاییز و زمستان ۱۳۸۴.
- [۶] نوروززاده، ی.؛ نظام آبادی پور، ح.؛ "ارائه ویژگی های دیداری جدید در تصویر دست نوشته فارسی برای تشخیص روحیات افراد"، مجله ماشین بینایی و پردازش تصویر، زمستان ۱۳۹۲.
- [۷] یغمایی، ف.؛ رضایی فر، و؛ سلامت منش، م.؛ عسگری، ع.؛ "پیش بینی خصوصیات روانشناختی افراد از روی دستخط های فارسی"، نخستین کنفرانس بین المللی پردازش خط و زبان فارسی، شهرریور ۱۳۹۱.
- [۸] نوروززاده، ی.؛ نظام آبادی پور، ح.؛ پوراآقاچوک، د.؛ "گرافولوژی دست نوشته به کمک پردازش تصویر و سیستم استنتاج فازی"، هشتمین کنفرانس سیستم های هوشمند، شهرریور ۱۳۸۶.
- [۹] صابری اناری، م.؛ احمدی، ع.؛ صابری اناری، م.؛ "تعیین میزان کنترل احساسات در افراد با کمک گرافولوژی و سیستم استنتاج فازی"، پانزدهمین کنفرانس دانشجویی مهندسی برق ایران، شهرریور ۱۳۹۱.
- [۱۰] مارنات، گروث، (۱۳۸۰)، "راهنمای سنجش روانی"، ترجمه پاشا شریفی و محمد رضا نیکخو، انتشارات رشد، تهران.
- [۱۱] واخوت، ل. دانشمند، (۱۳۵۷)، "ارزشیابی شخصیت"، انتشارات دانشگاه تهران.

[۱۲] م. میرزمانی، "بررسی مقیاسهای اعتباری فرم کوتاه پرسشنامه MMPI به زبان فارسی"، مجله روان شناسی و علوم تربیتی ۱۳۸۰ شماره ۲.

[۱۳] عربیان، ح.؛ قدس، و.؛ "یک پایگاه داده برای گرافولوژی"، بیستمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، مشهد، اسفند ۱۳۹۳.

[۱۴] دلاور، ع.؛ روش تحقیق در روانشناسی و علوم تربیتی، نشر ویرایش، بهار ۱۳۸۷.

[15] Tom Mitchell, McGraw-Hill, Machine Learning, 1977, Ch 3. Decision Tree Learning

[۱۶] میرزایی، ک.؛ سیستم های خبره و مهندسی دانش، ویراست هفتم پاییز ۱۳۹۱.

[۱۷] قدس، و.؛ کبیر، ا.؛ "استخراج ویژگی های حروف بر خط فارسی به منظور گروه بندی آن ها با استفاده از درخت تصمیم"، ششمین

کنفرانس ماشین بینایی و پردازش تصاویر ایران، ۵ و ۶ آبان ۱۳۸۹

[۱۸] یغمایی، ف.؛ سهرابی، م.؛ بیطرف، ح.؛ "تاثیر حالت روحی عجله

بر روی دستخط های زبان فارسی با استفاده از شبکه عصبی"،

هشتمین کنفرانس ماشین بینایی و پردازش تصویر ایران، دانشگاه

زنجان، ۱۳۹۲



وحيد قدس مدرک کارشناسی و کارشناسی

ارشد خود را در رشته مهندسی برق

الکترونیک به ترتیب از دانشگاه خواجه

نصیرالدین طوسی و دانشگاه سمنان در

سال های ۱۳۸۲ و ۱۳۸۴ دریافت کرد.

همچنین در سال ۱۳۹۱ موفق به اخذ مدرک

دکتری تخصصی در رشته مهندسی برق

الکترونیک از دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران شد. وی

در حال حاضر استادیار گروه مهندسی برق دانشگاه آزاد اسلامی واحد

سمنان است. زمینه های پژوهشی مورد علاقه وی پردازش سیگنال،

پردازش تصویر، پردازش گفتار، بازشناسی دستنوشته و روش های بهینه

سازی ابتکاری است.

حوریه عربیان مدرک کارشناسی خود را در

سال ۱۳۷۸ در رشته مهندسی برق

الکترونیک از دانشگاه سمنان و مدرک

کارشناسی ارشد خود را در رشته مهندسی

مکاترونیک در سال ۱۳۹۳ از دانشگاه آزاد

اسلامی پردیس علوم و تحقیقات سمنان

دریافت کرد. وی هم اکنون مدیر آموزشی-

پژوهشی مرکز علمی کاربردی جهاد

دانشگاهی سمنان می باشد. زمینه های پژوهشی مورد علاقه او پردازش

تصویر و سیستم های خبره است.

