

## تعیین هویت نویسنده با استفاده از اختلاط خبره‌ها (ME) و به کارگیری الگوریتم PCA جهت کاهش ویژگی‌ها در متون دستنویس فارسی

وحید زارعی<sup>۱</sup>، امیدرضا معروضی<sup>۲</sup>، حسین خسروی<sup>۳</sup>، علی قنبری سرخی<sup>۴</sup>

۱- دانشجوی کارشناسی ارشد الکترونیک دانشگاه صنعتی شاهرود، vahid\_zareii@yahoo.com

۲- عضو هیات علمی گروه الکترونیک دانشکده برق دانشگاه صنعتی شاهرود، Ormzi@yahoo.com

۳- عضو هیات علمی گروه الکترونیک دانشکده برق دانشگاه صنعتی شاهرود، Hosseinkhosravi@gmail.com

۴- دانشجوی کارشناسی ارشد مهندسی کامپیوتر، ali.ghanbari@gmail.com

### چکیده

در این مقاله روشی نو برای تعیین هویت نویسنده بر اساس متن دستنویس فارسی بصورت برون خط و مستقل از متن بر مبنای اختلاط خبره‌ها ارائه شده است. در روش پیشنهادی از فیلترهای گابور، سوبل و روبرتز برای استخراج ویژگی استفاده شده و سپس با استفاده از روش تجزیه و تحلیل مولفه‌های اصلی PCA و با کمک الگوریتم نزدیکترین همسایه کاهش ویژگی انجام گرفته است. برای ارزیابی دقت تشخیص به کمک ویژگی‌های استخراج شده از سه شبکه عصبی MLP<sup>۲</sup> جهت دسته بندی استفاده شده است که نتایج آنها را با هم ترکیب کرده ایم. جهت آموزش شبکه عصبی از گرادیان نزولی و روش پس انتشار خطا بهره گرفته ایم. نتایج نشان دهنده بهبود ۱۰٫۵ درصدی (۹۷٫۷۵ درصد در برابر ۸۷٫۲۵ درصد) در مقایسه با تحقیق صورت گرفته بر روی همین پایگاه داده و همچنین رسیدن به زمان آموزش ۷۱٫۷۳ و ۱۱۱٫۲۱ ثانیه به ترتیب برای ویژگی‌های استخراج شده با روش سوبل و روبرتز است که بیانگر عملکرد بهتری در مرحله طبقه بندی از نظر دقت و زمان کمتری در مرحله آموزش نسبت به تحقیقات مشابه قبلی در زمینه تعیین هویت نویسنده از دستخط فارسی می باشد.

### کلمات کلیدی

اختلاط خبره‌ها، تعیین هویت نویسنده، دستنوشته، شبکه عصبی، کاهش ویژگی، PCA

ثابت ماندن متن نیست و با هر متنی می توان هویت نویسنده را تعیین نمود.

#### ۱- مقدمه

روش دوم، روش برخط (online) است که در این روش علاوه بر ویژگی‌های برون خط از اطلاعات دینامیکی مثل فشار قلم، ترتیب نوشتن، سرعت نوشتن، فرم ضربه‌های قلم و ... نیز استفاده می شود. هر چند دقت این روش بیشتر است ولی بخاطر نیاز به سخت افزار ویژه کاربرد آنها نسبت به روش برون-خط کمتر است.

در زمینه تشخیص هویت نویسنده از روی دستنوشته فارسی تحقیقات زیادی انجام نشده است. از کارهایی که در این زمینه انجام شده می توان به موارد زیر اشاره کرد:

الدمور و همکارش از اطلاعات لبه جهت تشخیص هویت نویسنده در زبان عربی که شباهت زیادی به دستنویس فارسی دارد استفاده نموده اند [۷]. شهابی نژاد فیلترهای گابور را برای استخراج ویژگیها از متن دستنویس فارسی به کار گرفت [۲]. صادقی و همکارش از ویژگیهای مبتنی بر گرادیان برای آموزش شبکه عصبی چندلایه استفاده کرده اند [۸]. دهقان و همکاران یک پایگاه داده شامل دستنویس ۵۰ نویسنده تهیه کرده اند که از هر نویسنده ۱۰ نمونه از متن دلخواه در آن وجود دارد. در تحقیق ایشان از فیلترهای سوبل و روبرتز جهت استخراج

تعیین و تایید هویت بر اساس ویژگیهای بیومتریک، از زمینه‌های تحقیقاتی فعال است که در دهه اخیر مورد توجه فراوان قرار گرفته است [۱]. ویژگی‌های بیومتریک انسان به دو دسته ویژگی‌های فیزیولوژی و رفتاری تقسیم می گردند، که دستخط افراد جزء ویژگی‌های رفتاری محسوب می گردد. مطالعات نشان می دهد که افراد مختلف دستخطهای متفاوتی دارند [۶]. برای این منظور از دو روش کلی برای شناسایی استفاده می شود: روش برون خط و روش برخط.

در روش برون خط (off line) فقط تصویر متن دستنویس در دسترس است و ویژگی‌ها با توجه به ساختار کلمه‌ها و نویسه‌ها استخراج می گردند. در این روش تعداد زیادی از اطلاعات دینامیکی مربوط به طرز نوشتن افراد، که در روش برخط (on line) در دسترس است از دست می رود و به همین دلیل، نسبت به روش برخط دقت کمتری دارد [۶]. همچنین روش برون خط، خود به دو گروه کلی مستقل از متن و وابسته به متن تقسیم بندی می شود. در روش وابسته به متن باید متنی که توسط نویسنده ارائه می گردد ثابت باشد و در روشهای مستقل از متن، نیازی به