

استفاده از الگوریتم‌های داده کاوی در تشخیص اولیه بیماری پارکینسون در داده‌های متوازن

ناهید علاقه‌بند حسینی^۱، محمد مهدی حسینی^۲

۱- دانشجوی کارشناسی ارشد مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد شاهرود، ایران.

۲- استادیار، عضو هیات علمی گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد شاهرود، ایران.

چکیده: بیماری پارکینسون دارای تشخیص پزشکی بسیار مشکل و هزینه بر است و از آنجا که این بیماری در جوامع بشری بسیار شایع است، هر روز محققان در پی این هستند تا یک راه حل برای تشخیص زود هنگام این بیماری بیابند. تشخیص زود هنگام این بیماری نیاز به یک روش تشخیص صحیح و قابل اعتماد دارد که به پزشکان اجازه می‌دهد که آنرا از سایر بیماری‌های مشابه تمیز دهند. بنابراین، یافتن یک روش تشخیص صحیح و موثر و همچنین عوامل خطر در بروز این بیماری، بسیار با اهمیت است. بطور کلی ترکیب طبقه‌بندها یک زمینه تحقیقاتی جدید در مبحث یادگیری ماشین و تشخیص الگو می‌باشد. روش‌های ترکیبی به این صورت است که مجموعه‌ای از طبقه‌بندها را با داده‌های آموزشی ایجاد کرده و میزان صحت را با انجام عملیات رای گیری بر روی نتایج آن‌ها بدست می‌آورند. از طرفی با وجود داده‌های نامتوازن و یادگیری در مجموعه داده‌های نامتوازن جایی که نمونه‌های طبقه اکثریت خیلی بیشتر از بقیه است، چالش مهمی در یادگیری ماشین است زیرا الگوریتم‌های قدیمی یادگیری ماشین، ممکن است به سمت طبقه اکثریت متمایل شوند و این مسئله صحت پیش‌بینی را در طبقه اقلیت پایین می‌آورد. در این مقاله روش زیرنمونه برداری تصادفی را بعد از مقایسه آن با روش‌های دیگر نمونه برداری مانند بیش‌نمونه برداری تصادفی و EasyEnsemble و ModifiedBagging، برای نمونه برداری مجموعه آموزش استفاده شد و سپس نتایج را با معیارهای Precision و Recall و معیار F و معیار G ارزیابی نموده تا توان پیش‌بینی طبقه‌بندها در مقابل داده‌های نامتوازن افزایش پیدا کند. در این مقاله روشی را بر مبنای استخراج ویژگی بر مبنای طبقه‌بند ترکیبی شد تا داده‌های آموزش را برای طبقه‌بند پایه ایجاد شود.

کلمات کلیدی: بیماری پارکینسون، طبقه بندی، ماشین بردار پشتیبان، طبقه بند ترکیبی