



## **A Novel Technique for Parkinson's Diagnosis: EEG-Based Recurrence Quantification and Machine Learning**

**Alireza Talesh Jafadideh<sup>1</sup> / Zarei, Asghar<sup>2\*</sup>**

<sup>1</sup> - Assistant Professor, School of Engineering Science, College of Engineering, University of Tehran, Tehran, Iran

<sup>2</sup> - Assistant Professor, Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran.

### **ARTICLE INFO**

DOI: 10.22041/ijbme.2025.2052278.1956

Received: 1/2/2025

Revised: 27/4/2025

Accepted: 28/4/2025

### **KEY WORDS**

*Parkinson's disease*  
*EEG signal*  
*Recurrence Quantification Analysis (RQA)*  
*Machine Learning (ML)*

### **ABSTRACT**

Parkinson's disease (PD) occupies the second position among neurodegenerative disorders in terms of global prevalence, its defining characteristic being a significant deficiency of dopamine in the central nervous system. Consequently, diagnosing PD poses significant challenges, often involving a lengthy process. This has driven extensive research efforts to identify reliable biomarkers for PD. One approach to identifying PD involves analyzing the characteristics of EEG signals. EEG records brain activity by measuring electrical signals from electrodes placed on the scalp. The emergence of Artificial Intelligence (AI) has enabled the integration of EEG signal features into machine learning (ML) algorithms, facilitating automated diagnosis of neurological diseases. These findings suggest that EEG signals hold significant potential as biomarkers for distinguishing between individuals with PD and healthy controls. This research explores the feasibility of using features extracted from EEG signals through Recurrence Quantification Analysis (RQA) as potential biomarkers for Parkinson's Disease (PD). Utilizing publicly available EEG data from The Patient Repository for EEG Data + Computational Tools (PRED + CT), we analyzed recordings from PD patients who underwent repeated auditory stimulation tests. Different machine learning models, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Gentle Boosting (GentleBoost), Quadratic Discriminant Analysis (QDA), and Multilayer Perceptron (MLP), were used in both multiclass and binary classification scenarios. The performance of the aforementioned models was evaluated using a 10-fold cross-validation method. The proposed method achieved an average accuracy of 99.32% in the binary classification scenario (distinguishing between Parkinson's disease patients and healthy individuals) and an average accuracy of 98.30% in the multiclass classification using the KNN classifier. The results suggest that the RQA-based features extracted from electroencephalographic signals show promising potential as biomarkers for Parkinson's disease.



## روشی نوین در تشخیص بیماری پارکینسون: کمی‌سازی بازگشت مبتنی بر سیگنال‌های الکتروانسفالوگرافی و یادگیری ماشین

طالب جفادیده، علیرضا<sup>۱</sup> / زارعی، اصغر<sup>۲\*</sup>

<sup>۱</sup> - استادیار، گروه مهندسی بین رشته‌ای، دانشکده علوم مهندسی، دانشکده‌گان فنی، دانشگاه تهران، تهران، ایران

<sup>۲</sup> - استادیار، گروه مهندسی پزشکی، دانشکده مهندسی پزشکی، دانشگاه صنعتی سهند، تبریز، ایران

### مشخصات مقاله

شناسه‌ی دیجیتال: 10.22041/ijbme.2025.2052278.1956

پذیرش: ۱۴۰۴/۲/۸

بازنگری: ۱۴۰۴/۲/۷

ثبت در سامانه: ۱۴۰۳/۱۱/۱۳

### چکیده

### واژه‌های کلیدی

بیماری پارکینسون از نظر شیوع جهانی، دومین بیماری شایع مخرب عصبی محسوب می‌شود و مشخصه اصلی آن کمبود قابل توجه دوپامین در سیستم عصبی مرکزی است. تشخیص پارکینسون با چالش‌های قابل توجهی همراه بوده و اغلب شامل فرآیندی طولانی است. این امر باعث شده است تا تلاش‌های تحقیقاتی گسترده‌ای برای شناسایی زیست‌نشانه‌های قابل اعتماد برای پارکینسون انجام شود. یکی از روش‌های شناسایی پارکینسون پردازش سیگنال‌های الکتروانسفالوگرافی است. الکتروانسفالوگرافی فعالیت مغز را با اندازه‌گیری سیگنال‌های الکتریکی از الکترودهایی که روی جمجمه قرار می‌گیرند، ثبت می‌کند. ظهور هوش مصنوعی امکان ادغام ویژگی‌های سیگنال الکتروانسفالوگرافی در الگوریتم‌های یادگیری ماشین را فراهم کرده و تشخیص خودکار بیماری‌های عصبی را تسهیل می‌کند. یافته‌ها نشان می‌دهد که سیگنال‌های الکتروانسفالوگرافی پتانسیل قابل توجهی به‌عنوان زیست‌نشانه‌گر برای تمایز بین افراد مبتلا به پارکینسون و افراد سالم دارند. این مقاله امکان استفاده از ویژگی‌های استخراج شده از سیگنال‌های الکتروانسفالوگرافی از طریق ارزیابی کمی بازگشت را به‌عنوان زیست‌نشانه‌های بالقوه برای بیماری پارکینسون بررسی می‌کند. با استفاده از داده‌های الکتروانسفالوگرافی که به‌صورت عمومی از طریق (PRED + CT) در دسترس است، ما ثبت‌های بیماران پارکینسون را که تحت آزمایشات تحریک شنوایی مکرر قرار گرفته بودند، تجزیه و تحلیل کردیم. از مدل‌های یادگیری ماشین شامل: ماشین بردار پشتیبان،  $K$ -نزدیک‌ترین همسایه (KNN)، درخت تصمیم، جنگل تصادفی، روش GentleBoost، تحلیل تفکیکی درجه دوم (QDA) و پرسپترون چندلایه در دو حالت طبقه‌بندی چند کلاسه و دو کلاسه استفاده شد. عملکرد مدل‌های مذکور با استفاده از روش اعتبارسنجی متقابل ۱۰ فولد ارزیابی شد. روش پیشنهادی به میانگین صحت ۹۹/۳۲٪ در طبقه‌بندی دو کلاسه (تمایز بین بیماران پارکینسون و افراد سالم) و میانگین صحت ۹۸/۳۰٪ در طبقه‌بندی چندکلاسه با استفاده از دسته‌بندی کننده KNN دست یافت. نتایج نشان می‌دهند که ویژگی‌های کمی‌سازی بازگشتی استخراج شده از سیگنال‌های الکتروانسفالوگرافی نویدبخش به‌عنوان زیست‌نشانه‌های بالقوه برای پارکینسون هستند.

بیماری پارکینسون  
سیگنال الکتروانسفالوگرافی  
تحلیل کمی‌سازی بازگشتی  
یادگیری ماشین

## ۱- مقدمه

بیماری پارکینسون<sup>۱</sup> (PD) یک اختلال پیچیده و تحلیل‌برنده عصبی است که برای تشخیص دقیق و پایش آن به نشانگرهای زیستی قابل‌اعتماد نیاز است [۱-۳]. ویژگی‌های پیش‌رونده بیماری پارکینسون که شامل علائم حرکتی و غیرحرکتی می‌شود، اهمیت تشخیص زودهنگام را برای مداخله و مدیریت مؤثر آن برجسته می‌کند [۴]. در حالی که روش‌های سنتی تشخیص به شدت به مشاهدات بالینی وابسته هستند، ظهور رویکردهای مبتنی بر نشانگرهای زیستی مسیره‌های جدیدی را برای تشخیص عینی و دقیق گشوده است. پژوهش‌های اخیر انواع مختلفی از نشانگرهای زیستی را بررسی کرده‌اند که شامل سیگنال‌های صوتی، دست‌خط، الکتروانسفالوگرافی<sup>۲</sup> (EEG) و الگوهای راه رفتن است که با رویکردهای یادگیری ماشین برای تشخیص بیماری پارکینسون ترکیب شده‌اند [۵]. در میان این رویکردها، سیگنال‌های EEG به‌عنوان ابزاری غیرتهاجمی برای درک تغییرات عصبی مرتبط با بیماری پارکینسون امیدبخش ظاهر شده‌اند. مطالعات جدید نشان داده‌اند که ترکیب تحلیل سیگنال EEG با روش‌های پیشرفته یادگیری ماشین می‌تواند بینش‌های قابل‌توجهی درباره ویژگی‌های نوروفیزیولوژیکی مرتبط با این بیماری ارائه دهد [۶-۸]. بررسی مطالعات اخیر نشان می‌دهد که از سال ۲۰۰۴ تا ۲۰۲۳، پژوهش‌های مبتنی بر EEG در زمینه بیماری پارکینسون افزایش چشمگیری داشته‌اند [۹]؛ این موضوع بیانگر درک روزافزون اهمیت این روش در تشخیص و بررسی این بیماری است.

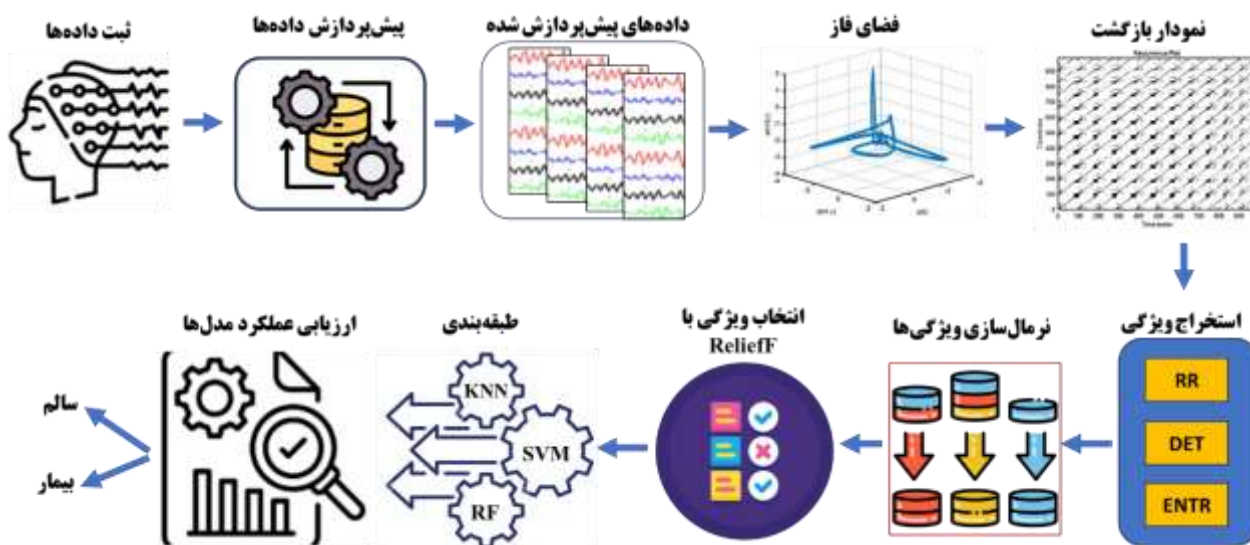
در سال‌های اخیر، استفاده از سیگنال‌های EEG برای شناسایی و تشخیص بیماری پارکینسون با بهبودها و دستاوردهای قابل‌توجهی همراه بوده است. به‌عنوان مثال، کوئلپو و همکاران استفاده از سیگنال‌های EEG و ویژگی‌های Hjorth را به‌عنوان نشانگرهای زیستی برای تشخیص PD بررسی کردند. برای طبقه‌بندی، آن‌ها عملکرد طبقه‌بندهای مختلف از جمله ماشین بردار پشتیبان<sup>۳</sup> (SVM)، K-نزدیک‌ترین همسایگی<sup>۴</sup> (KNN) و جنگل تصادفی<sup>۵</sup> (RF) را با استفاده از روش اعتبارسنجی متقابل ۵ فولد ارزیابی کردند. نتایج آن‌ها نشان داد که طبقه‌بند SVM با صحت ۸۹/۵۶٪ در تفکیک بیماران پارکینسون از افراد سالم بهترین عملکرد را داشت [۷]. اوه و همکاران به بررسی کاربرد

یک شبکه عصبی پیچشی<sup>۶</sup> از پیش آموزش دیده برای شناسایی موارد بیماری پارکینسون پرداختند. مدل توسعه‌یافته آن‌ها عملکرد قابل‌توجهی داشت و به صحت ۸۸/۲۵٪ دست یافت [۱۰]. لطیف‌اغلو و همکاران از روش‌های تجزیه حالت تجربی<sup>۷</sup>، تجزیه حالت متغیر<sup>۸</sup> و فیلتر ترتیب وولد-کالمن<sup>۹</sup> برای استخراج مجموعه‌ای از ویژگی‌ها از زیر باندهای EEG استفاده کردند که این ویژگی‌ها در مدل‌های مختلف یادگیری ماشین به کار گرفته شد. نتایج آن‌ها نشان داد که رویکرد فیلتر ترتیب وولد-کالمن مؤثرترین روش بود [۱۱]. کوچ و همکاران ویژگی‌های آماری و توان باندهای فرکانسی استخراج شده از سیگنال‌های EEG ۴۰ شرکت‌کننده را تحلیل کردند و این یافته‌ها را با متغیرهای جمعیت‌شناختی در یک چارچوب طبقه‌بندی جنگل تصادفی ترکیب کردند. آن‌ها با استفاده از نتایج به‌دست‌آمده از این رویکرد خودکار، مدلی توسعه دادند که به صحت ۹۱٪ دست یافت [۱۲]. الووی و همکاران پارامترهای Hjorth شامل فعالیت (Activity)، تحرک (Mobility) و پیچیدگی (Complexity) را از سیگنال‌های EEG استخراج کردند. آن‌ها رفتارهای همبستگی و خوشه‌بندی را بررسی کرده و الگوهای متمایزی را در داده‌های بیماران پارکینسون نسبت به گروه کنترل شناسایی کردند [۱۳]. در مطالعه‌ای دیگر، دی الیویرا و همکاران عملکرد ۹ طبقه‌بند را برای شناسایی بیماری پارکینسون بر اساس سیگنال EEG ثبت‌شده در حین یک تکلیف بصری ارزیابی کردند. ورودی‌ها شامل مقادیر همبستگی جهت‌دار<sup>۱۰</sup> بین جفت الکترودها، با تمرکز بر نواحی مغزی مرتبط با پردازش بصری و با استفاده از تنها ۱۰ الکتروود بود [۱۴]. کاواناق و همکاران مجموعه داده (PRED + CT) برای شناسایی بیماران پارکینسون از طریق تحلیل پتانسیل‌های مرتبط با رویداد استفاده کردند. آن‌ها از یک طبقه‌بند SVM بهره بردند که ۵۰ ضریب اولیه تبدیل فوریه سریع سیگنال در پاسخ به یک محرک را به‌عنوان ویژگی‌های ورودی به کار گرفت. این مطالعه به صحت حداکثری ۸۲٪ در تفکیک بین افراد سالم و مبتلایان به بیماری پارکینسون دست یافت [۱۵].

تحلیل کمی‌سازی بازگشتی<sup>۱۱</sup> (RQA)، به‌عنوان ابزاری بسیار جذاب برای ارزیابی سیستم‌های دینامیکی غیرخطی، به یک

<sup>۶</sup> Empirical mode decomposition<sup>۸</sup> Variational mode decomposition<sup>۹</sup> Vold-Kalman order filtering<sup>۱۰</sup> Directed coherence values<sup>۱۱</sup> Recurrence Quantification Analysis<sup>۱</sup> Parkinson's Disease<sup>۲</sup> Electroencephalography<sup>۳</sup> Support Vector Machine<sup>۴</sup> K-Nearest Neighbors<sup>۵</sup> Random Forest<sup>۶</sup> Convolutional neural network

روش قدرتمند برای استخراج ویژگی‌های معنادار از سیگنال‌های EEG تبدیل شده است. اثربخشی منحصربه‌فرد این روش در



شکل ۱- روندنمای روش پیشنهادی برای طبقه‌بندی خودکار بیماری پارکینسون.

با بهره‌گیری از ترکیب ویژگی‌های RQA، سیگنال‌های EEG و الگوریتم‌های یادگیری ماشین، هدف ما توسعه رویکردی دقیق و کارآمد برای تشخیص و پایش بیماری پارکینسون است که به گسترش دانش در این حوزه مهم کمک خواهد کرد.

ساختار این مقاله به شرح زیر سازمان‌دهی شده است: **بخش دوم** توضیح دقیقی از مجموعه داده و روش‌های پیشنهادی ارائه می‌دهد و نمای کلی از داده‌ها و چارچوب تحلیلی به‌کاررفته در پژوهش را بیان می‌کند. **بخش سوم** به بررسی نتایج و پیامدهای آن‌ها می‌پردازد و تحلیلی عمیق از یافته‌ها ارائه می‌دهد. در نهایت، **بخش چهارم** با خلاصه کردن نتایج کلیدی و تأکید بر اهمیت آن‌ها برای این تحقیق، مقاله را به پایان می‌رساند.

## ۲- مواد و روش‌ها

این بخش به تشریح روش‌شناسی به‌کاررفته برای جمع‌آوری داده‌ها و فرآیندهای تحلیلی اعمال‌شده بر ویژگی‌هایی که در طبقه‌بندی بیماران مبتلا به پارکینسون استفاده شده‌اند، می‌پردازد. همانطور که در بخش قبلی ذکر شد، استفاده از ویژگی‌های استخراج‌شده از RQA همراه با الگوریتم‌های یادگیری ماشین، چارچوب مناسبی برای توسعه نشانگرهای زیستی مشخص برای بیماری پارکینسون (PD) فراهم می‌کند. روندنمای روش پیشنهادی در این مقاله در شکل ۱ ارائه شده است. هدف این پژوهش بررسی ویژگی‌های مبتنی بر RQA استخراج‌شده از سیگنال‌های EEG به‌عنوان نشانگرهای زیستی

شناسایی تغییرات جزئی در پویایی‌های مغزی که ممکن است با روش‌های تحلیلی سنتی قابل شناسایی نباشند، مورد توجه قرار گرفته است [۱۶].

روش RQA با بهره‌گیری از نمودارهای بازگشتی، امکان شناسایی ساختارهای پویای پیچیده و وابستگی‌های زمانی پنهان در سیگنال EEG را فراهم می‌سازد. این ویژگی به‌ویژه در تحلیل سیگنال‌های غیرایستا و آشوبناک، همانند سیگنال EEG در بیماران پارکینسونی، بسیار حائز اهمیت است برخلاف روش‌های آنتروپی که بر پیچیدگی کلی تمرکز دارند و یا ابعاد فراکتال که بیشتر به شباهت‌های مقیاسی می‌پردازند، RQA قادر به شناسایی الگوهای بازگشتی بوده و اطلاعات دقیقی در مورد دینامیک‌های گذرا و وابستگی‌های زمانی فراهم می‌کند [۱۷، ۱۸]. به‌طور خاص، معیارهایی مانند Determinism، Laminarity و Trapping Time که توسط RQA استخراج می‌شوند، شاخص‌هایی برای سنجش میزان پیش‌بینی‌پذیری، پایداری و مدت ماندگاری در وضعیت‌های خاص مغزی هستند. ویژگی‌هایی نظیر Trapping Time و Laminarity شاخص‌هایی از پایداری دینامیکی فعالیت عصبی هستند که در طیف وسیعی از اختلالات نورولوژیکی، از جمله بیماری‌های تحلیل‌برنده عصبی نظیر پارکینسون، دچار نوسانات قابل توجهی می‌گردند [۱۹]. همچنین باید اشاره کرد که RQA پیش‌تر در زمینه‌های متعددی از جمله تحلیل مراحل خواب، بررسی عمق بیهوشی و تشخیص صرع با موفقیت به‌کاررفته است، که این مسئله نیز نشان‌دهنده پایداری و کارایی این روش در تحلیل سیگنال‌های نورودینامیکی می‌باشد.

نداشت، حذف شدند و در نهایت، ۶۰ کانال برای تحلیل باقی ماند. سیگنال‌های EEG فیلترشده به بازه‌های زمانی تقسیم شدند که این بازه‌ها شامل ۲ ثانیه قبل و بعد از تحریک شنیداری بودند. در این مطالعه کل پنجره ۴ ثانیه (شامل ۲۰۰۰ نمونه زمانی) برای پردازش‌های بعدی استفاده شده است. تمامی کانال‌ها و بازه‌هایی که دارای آرتیفکت (نویز و اختلال) بودند، معیوب در نظر گرفته شده و حذف شدند. در مقابل، آرتیفکت‌های چشمی شامل حرکات چشم و پلک زدن از طریق تحلیل مؤلفه‌های مستقل (ICA) اصلاح شدند. اطلاعات بیشتر درباره این مجموعه داده در مرجع [۱۵] موجود است.

## ۲-۲- محرک‌های شنوایی

در الگوی Oddball معمولاً همبستگی‌های عصبی شناخت و توجه از طریق سیگنال‌های EEG بررسی می‌شود. در این آزمایش، به شرکت‌کنندگان دو دسته محرک ارائه شد: محرک استاندارد و محرک هدف. شرکت‌کنندگان هنگام مشاهده محرک هدف باید به صورت ذهنی آن را می‌شمردند یا دکمه‌ای را فشار می‌دادند و در عین حال محرک‌های استاندارد را نادیده می‌گرفتند. این مطالعه از نسخه‌ای تغییر یافته از این الگو استفاده کرد که به آن وظیفه شنیداری سه‌تحریکی Oddball گفته می‌شود [۱۵]. این نسخه شامل سه نوع محرک شنیداری مختلف بود: محرک استاندارد، محرک هدف و محرک انحرافی. هر آزمون به‌طور خلاصه شامل سه صدای مختلف بود و به شرکت‌کنندگان دستور داده شد که فقط محرک‌های هدف را بشمارند. محرک استاندارد یک صدای ۴۴۰ هرتزی بود که در ۷۰٪ از آزمون‌ها ارائه می‌شد، محرک هدف یک صدای ۶۶۰ هرتزی بود که در ۱۵٪ از آزمون‌ها ظاهر می‌شد و محرک انحرافی یک صدای طبیعی بود که آن هم در ۱۵٪ از آزمون‌ها ارائه می‌شد. این آزمایش شامل دو بلوک بود که هر کدام شامل ۱۰۰ محرک بودند. در طول آزمایش، از شرکت‌کنندگان خواسته شد تعداد محرک‌های هدف را که شنیده‌اند گزارش دهند. پاسخ‌های آن‌ها در شمارش محرک‌های هدف با انتظارات منطقی مرتبط با این تکلیف مطابقت داشت و درستی درک آن‌ها از دستورالعمل‌ها را تأیید می‌کرد. کل آزمایش، شامل انجام تکالیف شنیداری و ثبت داده‌ها، به‌طور میانگین حدود ۱۲ دقیقه طول کشید.

## ۲-۳- استخراج ویژگی‌ها

روش RQA یک روش تحلیلی مفید است که برای کمی‌سازی رفتار بازگشتی حالت‌ها در سیستم‌های دینامیکی طراحی شده است [۱۶]. محاسبات RQA شامل چندین مرحله ریاضی

برای بیماری پارکینسون است. مقاله حاضر با بهره‌گیری از پیشرفت‌های اخیر، رویکردی نوین را جهت استخراج و طبقه‌بندی ویژگی‌ها ارائه می‌دهد. علاوه بر این، در این مطالعه، تمرکز ما بر انتخاب بهترین ویژگی‌های مبتنی بر RQA است که با استفاده از الگوریتم ReliefF انتخاب می‌شوند. به‌منظور ارزیابی میزان پایایی نتایج حاصل از رویکرد پیشنهادی، عملکرد چندین طبقه‌بند از طریق روش اعتبارسنجی متقابل ۱۰-فولد بررسی شده است. علاوه بر این، عملکرد طبقه‌بندها با استفاده از چندین معیار کمی به‌طور دقیق بررسی می‌شود.

## ۲-۱- داده مورد استفاده

پایگاه داده مورد استفاده در این مطالعه از مخزن عمومی و دسترس PRED+CT استخراج شده است که داده‌های EEG را فراهم می‌کند و ابزارهای تحلیلی برای ارزیابی اختلالات مغزی ارائه می‌دهد [۱۵]. این پایگاه داده شامل بیماران با تشخیص بالینی بیماری پارکینسون بود و گروه کنترل سالم متشکل از ۲۵ نفر با تطابق سنی و جنسیتی با بیماران داشت. تمامی شرکت‌کنندگان در یک وظیفه شنیداری شامل سه تحریک Oddball شرکت کردند که جزئیات آن در زیربخش‌های بعدی توضیح داده خواهد شد.

گروه بیماران مبتلا به پارکینسون در دو مرحله مورد آزمایش قرار گرفتند: مرحله‌ای با مصرف معمول داروی دوپامینرژیک و مرحله‌ای دیگر که در آن حداقل ۱۵ ساعت قبل از جلسه ثبت داده‌ها بدون دارو بودند. تمامی افراد شرکت‌کننده در مطالعه با استفاده از آزمون بررسی وضعیت ذهنی کوتاه<sup>۱</sup> (MMSE) ارزیابی شدند. بیماران مبتلا به پارکینسون و گروه کنترل از نظر سطح تحصیلات قابل مقایسه بودند و بر اساس نتایج MMSE هیچ‌کدام از آن‌ها دچار زوال عقل متوسط یا شدید نبودند. بیماران در هر دو مرحله با دارو و بدون دارو مورد ارزیابی قرار گرفتند. این ارزیابی شامل امتیازهای حاصل از MMSE و مقیاس یکپارچه ارزیابی بیماری پارکینسون<sup>۲</sup> (UPDRS) بود و همچنین داده‌هایی درباره میزان دوز لوودوپا، مدت زمان ابتلا به بیماری و جزئیات جمعیت‌شناختی را نیز در بر می‌گرفت که از منبع [۱۵] بازبینی شده است.

سیگنال EEG به‌صورت پیوسته با استفاده از ۶۴ الکتروود AgCl ثبت شد. کانال AFz به‌عنوان مرجع مورد استفاده قرار گرفت. نرخ نمونه‌برداری بر روی ۵۰۰ هرتز تنظیم شده بود و سیگنال از طریق یک فیلتر میان‌گذر با بازه ۰/۱ تا ۱۰۰ هرتز فیلتر شد. کانال‌هایی که سیگنال آن‌ها با الگوی معمول EEG مطابقت

<sup>۱</sup> Unified PD Rating Scale

<sup>۲</sup> Mini-Mental State Examination

می‌کند و بینش‌های معناداری در مورد رفتار سیستم، از جمله پایداری، قابلیت پیش‌بینی و پیچیدگی ارائه می‌دهد [۲۰]. ویژگی‌های مبتنی بر آنتروپی مانند آنتروپی تقریبی (ApEn) و آنتروپی نمونه‌ای (SampEn)، معیارهایی غیرخطی برای سنجش میزان بی‌نظمی و پیچیدگی سیگنال‌های زمان‌گسسته هستند. ApEn احتمال باقی‌ماندن الگوهای مشابه در طول زمان را اندازه‌گیری می‌کند، در حالی که SampEn با کاهش سوگیری در داده‌های کوچک، نسخه بهبود یافته‌ای از ApEn محسوب می‌شود. این ویژگی‌ها به‌طور مؤثری برای تشخیص تغییرات پیچیدگی در سیگنال EEG در شرایط مختلف شناختی یا پاتولوژیک به کار می‌روند. همچنین به دلیل هزینه محاسباتی پایین‌تر، برای تحلیل‌های بلندمدت مناسب هستند [۲۱]. با این حال، در مقایسه با روش RQA، حساسیت بیشتری نسبت به انتخاب پارامترها دارند، مقاومت کمتری در برابر نویز نشان می‌دهند و وابستگی‌های زمانی بلندمدت را به‌خوبی RQA منعکس نمی‌کنند. در این مطالعه، علاوه بر ویژگی‌های مبتنی بر RQA، عملکرد ویژگی‌های آنتروپی نمونه‌ای و تقریبی در شناسایی بیماران مبتلا به بیماری پارکینسون مورد بررسی قرار می‌گیرد.

#### ۴-۲- نرمال‌سازی ویژگی‌ها

نرمال‌سازی ویژگی‌ها با استفاده از روش نمره Z یک تکنیک پیش‌پردازش حیاتی در یادگیری ماشین و تحلیل داده است که مزایای متعددی را ارائه می‌دهد. این روش داده‌ها را استاندارد می‌کند تا میانگین برابر با صفر و انحراف برابر با یک باشد و بدین ترتیب تأثیر واحدها یا مقیاس‌های متفاوت میان ویژگی‌ها را از بین می‌برد. در نتیجه، ویژگی‌هایی با بازه‌های مختلف تأثیر یکسانی بر فرآیند مدل‌سازی خواهند داشت. این امر به‌ویژه برای الگوریتم‌هایی که به فواصل یا گرادیان‌ها حساس هستند، مانند رگرسیون خطی، K-نزدیکترین همسایه، ماشین بردار پشتیبان (SVM) و شبکه‌های عصبی بسیار مفید است، چرا که عملکرد مدل را بهبود می‌بخشد. در مطالعه حاضر، تمامی ویژگی‌ها با استفاده از روش نمره Z نرمال‌سازی شده‌اند.

#### ۴-۲-۵- انتخاب ویژگی

انتخاب ویژگی نقش بسیار مهمی در مرحله پردازش داده‌ها برای مدل‌های یادگیری ماشین ایفا می‌کند و این فرآیند با

است که با بازسازی فضای فاز آغاز می‌شود و به محاسبه نهایی معیارهای مبتنی بر بازگشت ختم می‌شود. در ادامه، توضیح دقیقی از نحوه پیاده‌سازی ریاضی RQA ارائه شده است. اولین گام، بازسازی فضای فاز است. برای بازسازی فضای فاز از یک سری زمانی  $x(t)$ ، از تکنیکی به نام تعبیه با تأخیر زمانی<sup>۱</sup> استفاده می‌شود:

$$x_i = [x(i), x(i + \tau), x(i + 2\tau), \dots, x(i + (m - 1)\tau)] \quad (1)$$

که در آن  $x(i)$  بردار تعبیه شده را در فضای فاز در زمان  $i$  نشان می‌دهد. همچنین  $m$  و  $\tau$  به ترتیب نشان دهنده بعد تعبیه و تأخیر زمانی هستند. در مرحله دوم، ماتریس بازگشتی  $R_{i,j}$  محاسبه می‌شود تا مشخص شود که دو حالت  $x(i)$  و  $x(j)$  در فضای فاز به هم نزدیک هستند [۲۰]:

$$R_{i,j} = \theta(\varepsilon - \|x_i - x_j\|), \quad i, j = 1, \dots, N \quad (2)$$

که در آن  $\| \cdot \|$  معیار فاصله بین دو نقطه است. همچنین  $\varepsilon$  و  $\theta$  به ترتیب یک آستانه از پیش تعریف شده و تابع گام Heaviside هستند.

ماتریس بازگشتی  $R$  یک ماتریس باینری متقارن  $N \times N$  است که در آن  $R_{i,j} = 1$  نشان دهنده بازگشت و  $R_{i,j} = 0$  نشان دهنده عدم بازگشت هستند [۲۰]. در مرحله سوم، چندین معیار RQA با استفاده از ماتریس بازگشتی  $R$  محاسبه می‌شوند. نرخ بازگشت (RR) به‌عنوان نسبت نقاط بازگشت در ماتریس مشخص می‌شود:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j} \quad (3)$$

جبرگرای<sup>۲</sup> (DET) به‌عنوان نسبت نقاط بازگشتی که ساختارهای مورب به طول  $l \geq l_{min}$  را ایجاد می‌کنند به تعداد کل نقاط بازگشتی تعریف می‌شود:

$$DET = \frac{\sum_{l=l_{min}}^N IP(l)}{\sum_{l=1}^N IP(l)} \quad (4)$$

که در آن  $P(l)$  توزیع فرکانس مرتبط با خطوط مورب با طول  $l$  را نشان می‌دهد [۲۰]. آنتروپی شانون توزیع طول خط مورب<sup>۳</sup> (ENTR) را می‌توان با استفاده از رابطه (۵) محاسبه کرد:

$$ENTR = \sum_{l=l_{min}}^N P(l) \log(P(l)) \quad (5)$$

با محاسبه این ویژگی‌ها، تحلیل کمی بازگشت، پویایی‌های یک سیستم را بر اساس ویژگی‌های بازگشتی آن کمی‌سازی

<sup>۱</sup> Shannon entropy of the diagonal line lengths distribution

<sup>۱</sup> Time-delay embedding

<sup>۲</sup> Determinism

جدول ۱: مقادیر بهینه پارمترهای هر یک از طبقه‌بندها.

پارامترها	طبقه‌بند
'NumNeighbors'= 3, 'Distance'='cosine', 'DistanceWeight'='inverse', 'Standardize'= true	KNN
'Standardize'= true, 'KernelFunction'='RBF', 'BoxConstraint'= 10, 'KernelScale'=0.25	SVM (RBF)
'Standardize'= true, 'KernelFunction'='polynomial', 'PolynomialOrder'=3	SVM (polynomial)
'MaxNumSplits'= 705	DT
'NumLearningCycles' = 50	RandomForest
'NumLearningCycles' = 120, 'LearnRate' = 0.0017	GentleBoost
'DiscrimType' = 'quadratic'	QDA
regularization = 0.2, transferFcn = 'logsig', hiddenSizes = [30,10]	MLP

برای ارزیابی عملکرد طبقه‌بندهای مختلف، مجموعه داده ابتدا به‌طور تصادفی مخلوط و سپس با استفاده از روش اعتبارسنجی متقابل ۱۰ فولد به ده زیرمجموعه مجزا تقسیم شد. در این روش، یک زیرمجموعه به‌عنوان مجموعه آزمایش تعیین و سایر زیرمجموعه‌ها برای آموزش مدل استفاده می‌شوند. این فرآیند به‌صورت تکراری برای هر زیرمجموعه انجام می‌شود. برای ارزیابی اثربخشی مدل، از معیارهای کمی مختلف استفاده شد که شامل موارد زیر است: ماتریس درهم‌ریختگی، منحنی ROC، صحت (Accuracy)، دقت (Precision)، حساسیت (Sensitivity)، اختصاصیت (Specificity)، نمره  $F_1$  (F1-score).

ماتریس درهم‌ریختگی یکی از مهم‌ترین ابزارهای ارزیابی عملکرد مدل‌های طبقه‌بندی است که میزان صحت، حساسیت، اختصاصیت، دقت و نمره  $F_1$  پیش‌بینی‌های مدل را از طریق چهار مؤلفه مثبت واقعی<sup>۳</sup> (TP)، منفی واقعی<sup>۴</sup> (TN)، مثبت کاذب<sup>۵</sup> (FP) و منفی کاذب<sup>۶</sup> (FN) نمایش می‌دهد. منحنی ROC نشان‌دهنده تعادل بین نرخ مثبت واقعی (حساسیت) و نرخ مثبت کاذب (یک منهای اختصاصیت) در آستانه‌های مختلف طبقه‌بندی است. این منحنی ابزاری حیاتی برای ارزیابی توانایی یک مدل در تفکیک بین کلاس‌های مختلف محسوب می‌شود، در حالی که مساحت زیر منحنی ROC (AUC)<sup>۷</sup> یک معیار خلاصه و جامع از عملکرد کلی مدل ارائه می‌دهد.

علاوه بر موارد مذکور، در این مقاله از آزمون ویلکاکسون (دستور signrank در متلب) برای تحلیل آماری ویژگی‌های انتخاب شده از نظر ایجاد تفاوت معنی‌دار بین گروه‌های مختلف استفاده شد.

انتخاب مهم‌ترین ویژگی‌ها می‌تواند دقت و کارایی مدل را افزایش دهد. این کار با کاهش بُعد داده‌ها باعث کاهش پیچیدگی محاسباتی می‌شود و هم‌زمان با حذف ویژگی‌های نامربوط، از بیش‌برازش<sup>۱</sup> جلوگیری می‌کند. الگوریتم ReliefF یک الگوریتم کارآمد برای انتخاب ویژگی است که عمدتاً برای مجموعه داده‌های چندکلاسه و دارای نویز استفاده می‌شود [۲۱]. این روش با مقایسه نزدیک‌ترین همسایه‌های مشابه و نامشابه، قدرت تفکیک‌پذیری هر ویژگی را برآورد کرده و برای هر ویژگی یک امتیاز وزنی اختصاص می‌دهد. الگوریتم ReliefF نسبت به همبستگی‌های غیرخطی حساس است و تعاملات بین ویژگی‌ها را به‌خوبی شناسایی می‌کند، که این موضوع آن را به یک روش قدرتمند برای انتخاب ویژگی در مجموعه داده‌های پیچیده تبدیل می‌کند. در این مطالعه، استفاده از این الگوریتم باعث شناسایی ویژگی‌های کلیدی و حذف متغیرهای غیرمؤثر شده است و تأثیر مثبت آن بر عملکرد مدل را برجسته می‌کند.

## ۶-۲- طبقه‌بندی

در این پژوهش، سه وظیفه طبقه‌بندی دوتایی به‌منظور تمایز میان سه گروه انجام شد: گروه کنترل (CTRL)، افرادی با بیماری پارکینسون که دارو مصرف نمی‌کردند (PDOFF) و بیماران مبتلا به پارکینسون که تحت درمان دارویی بودند (PDON). همچنین، طبقه‌بندی سه کلاسه بین گروه‌های سالم (CTRL)، PDON و PDOFF انجام شد. در این پژوهش، به منظور انجام فرآیند دسته‌بندی، از طبقه‌بندهای متداول شامل KNN، SVM با توابع هسته‌ای مختلف، درخت تصمیم (DT)، جنگل تصادفی، الگوریتم تقویت تدریجی (GentleBoost)، تحلیل تفکیکی درجه دوم (QDA) و شبکه پرسپترون چندلایه (MLP) استفاده شده است. مقادیر بهینه پارمترهای هر یک از طبقه‌بندها از طریق فرآیند جستجوی شبکه‌ای به دست آمده و به منظور افزایش شفافیت پژوهش، در جدول ۱ ارائه گردیده است. لازم به ذکر است که مقادیر بهینه با هدف بهبود دقت طبقه‌بندی و جلوگیری از بیش‌برازش انتخاب شده‌اند.

<sup>۵</sup> False positive

<sup>۶</sup> False negative

<sup>۷</sup> Area Under Curve

<sup>۱</sup> Overfitting

<sup>۲</sup> Receiver Operating Characteristic

<sup>۳</sup> True positive

<sup>۴</sup> True negative

الگوریتم SVM با تابع هسته RBF با فراهم نمودن میانگین صحت ۹۸/۶۵٪ بهترین عملکرد طبقه‌بندی را ارائه داد. این نتایج نشان می‌دهد که مصرف دارو قبل از جمع‌آوری داده‌ها تأثیر قابل توجهی بر الگوهای سیگنال مغزی دارد و منجر به تفاوت‌های قابل اندازه‌گیری در ویژگی‌های سیگنال می‌شود.

جدول ۲- الکترودها و ویژگی‌های دارای بیشترین اطلاعات در سه حالت مختلف طبقه‌بندی.

PDOFF × PDON		CTRL × PDON		CTRL × PDOFF	
ویژگی	کانال	ویژگی	کانال	ویژگی	کانال
'DET'	'FC4'	'RR'	'Fp1'	'RR'	'Fp1'
'RR'	'FCz'	'DET'	'Fp1'	'DET'	'Fp1'
'DET'	'FCz'	'DET'	'FCz'	'RR'	'AF7'
'RR'	'C1'	'DET'	'F8'	'DET'	'AF7'
'RR'	'FC4'	'RR'	'F8'	'DET'	'T7'
'DET'	'FC6'	'RR'	'FCz'	'RR'	'T7'
'DET'	'CP5'	'RR'	'O2'	'DET'	'F8'
'RR'	'CP5'	'ENTR'	'Fp1'	'RR'	'F8'
'RR'	'FC6'	'DET'	'O2'	'ENTR'	'AF7'
'DET'	'C1'	'DET'	'AF7'	'ENTR'	'Fp1'
'DET'	'FC5'	'RR'	'AF7'	'RR'	'T8'
'ENTR'	'Fp2'	'DET'	'F1'	'DET'	'T8'
'ENTR'	'AF7'	'RR'	'Fp2'	'RR'	'TP7'
'RR'	'AFz'	'RR'	'F1'	'DET'	'F5'
'DET'	'AFz'	'RR'	'FC1'	'DET'	'TP7'
'RR'	'FC5'	'DET'	'FC1'	'DET'	'FT7'
'DET'	'F1'	'DET'	'FC5'	'RR'	'FC1'
'RR'	'F6'	'ENTR'	'AF7'	'DET'	'FC1'
'RR'	'F1'	'DET'	'F7'	'RR'	'F5'
'RR'	'C2'	'DET'	'Fp2'	'RR'	'AF4'

همچنین، نتایج به‌دست‌آمده از مقایسه سه روش تحلیل غیرخطی شامل تحلیل کمی بازگشتی، آنتروپی نمونه‌ای و آنتروپی تقریبی در جدول ۴، نشان‌دهنده‌ی عملکرد رقابتی و قابل اتکای RQA در طبقه‌بندی سیگنال‌های EEG مرتبط با بیماری پارکینسون است. در حالی که SampEn در برخی حالت‌ها دقت اندکی بالاتری ارائه داده است، روش RQA با ارائه دقتی بالا و انحراف معیار کمتر، پایداری قابل توجهی از خود نشان داده است. به‌ویژه در طبقه‌بندی بین گروه کنترل و بیماران در وضعیت بدون دارو (PDOFF)، RQA عملکردی نزدیک به SampEn و حتی برتر از ApEn داشته است. همچنین، توانایی RQA در شناسایی الگوهای بازگشتی و وابستگی‌های زمانی پیچیده، آن را به گزینه‌ای مناسب برای تحلیل‌های دقیق‌تر در شرایط پویای مغزی بدل می‌سازد. این نتایج مؤید آن است که RQA می‌تواند به عنوان ابزاری مؤثر در کنار سایر روش‌های غیرخطی برای شناسایی تغییرات پاتولوژیک در سیگنال EEG مورد استفاده قرار گیرد.

در ادامه ماتریس‌های درهم‌ریختگی فقط برای سه نوع طبقه‌بندی عملکرد مناسبی در سه حالت دسته‌بندی دوتایی

### ۳- نتایج و بحث

همان‌طور که در بخش‌های قبلی اشاره شد، این مطالعه یک روش خودکار برای شناسایی افراد مبتلا به بیماری پارکینسون (PDON یا PDOFF) با استفاده از رویکردهای یادگیری ماشین پیشنهاد داده است. روش پیشنهادی از ویژگی‌های مبتنی بر تحلیل کمی‌سازی بازگشتی برای اندازه‌گیری تفاوت‌های سیگنال‌های مغزی بین گروه سالم (CTRL) و افراد مبتلا به پارکینسون استفاده می‌کند. فرآیند استخراج ویژگی‌ها برای هر بخش از داده‌ها ۱۸۰ ویژگی تولید کرد (۶۰ کانال  $\times$  ۳ ویژگی: DET، ENTR و RR). در ادامه، الگوریتم ReliefF برای انتخاب بهترین زیرمجموعه از ویژگی‌ها به‌کار گرفته شد. در این مطالعه، ۵۰ ویژگی برتر از خروجی الگوریتم ReliefF به‌عنوان مجموعه ویژگی‌های بهینه انتخاب شدند. جدول ۲ خلاصه‌ای از جزئیات مربوط به ۲۰ ویژگی برتر را ارائه می‌دهد که شامل شناسه ویژگی و کانال انتخابی متناظر با آن است. همان‌طور که در این جدول نشان داده شده است، ویژگی‌های انتخاب‌شده عمدتاً از نواحی پیشانی (Frontal)، مرکزی (Central) و پیشانی-مرکزی (Fronto-Central) استخراج شده‌اند. این مشاهده نشان می‌دهد که سیگنال‌های این نواحی مغزی حاوی اطلاعات ارزشمندی برای تفکیک بین گروه CTRL و افراد مبتلا به پارکینسون هستند. مقدار آماره  $p$  برای ویژگی‌های انتخاب‌شده کمتر از ۰/۰۱ می‌باشد که نشان‌دهنده تفاوت معنادار آماری بین گروه‌های سالم و بیمار است.

نتایج طبقه‌بندی ویژگی‌های انتخاب‌شده با استفاده از الگوریتم‌های SVM با تابع هسته چندجمله‌ای و تابع پایه گوسی (RBF)، KNN، RF، درخت تصمیم، GentleBoost، QDA و MLP در جدول ۳ ارائه شده است. ارزیابی عملکرد این طبقه‌بندها شامل گزارش معیارهای مختلفی مانند صحت، حساسیت، اختصاصیت، مساحت زیر منحنی ROC، دقت و نمره  $F_1$  بود. نتایج نشان می‌دهد که در حالت طبقه‌بندی گروه CTRL در مقابل گروه PDOFF، طبقه‌بند KNN با میانگین دقت ۹۹/۳۲٪ بهترین عملکرد را داشت، در حالی که طبقه‌بند DT با میانگین دقت ۸۶/۹۸٪ بدترین عملکرد را نشان داد. در طبقه‌بندی بین گروه CTRL و PDON نیز طبقه‌بند KNN بار دیگر عملکرد برتری داشت و به میانگین صحت ۹۹/۰۴٪ دست یافت. قابل توجه است که طبقه‌بند SVM با تابع هسته RBF نیز عملکرد قابل مقایسه‌ای با میانگین صحت ۹۸/۹۳٪ داشت و به‌عنوان یک گزینه رقابتی برای این تکلیف طبقه‌بندی مطرح شد. در حالت طبقه‌بندی PDOFF در مقابل PDON،

کمتر است. این امر نشان می‌دهد که اگرچه KNN در تشخیص نمونه‌های مثبت عملکرد بهتری دارد، اما اختلاف آن با SVM از نظر FN چندان قابل توجه نیست. علاوه بر این، مقدار تشخیص‌های درست (TP+TN) در KNN بیشترین مقدار را در میان سه روش دارد، که نشان‌دهنده عملکرد کلی بهتر این طبقه‌بند در دسته‌بندی صحیح نمونه‌ها است. پس از KNN، مقدار TP+TN در SVM بیشتر از GentleBoost است، که تأیید می‌کند GentleBoost در مقایسه با دو روش دیگر عملکرد ضعیف‌تری دارد.

داشتند، در شکل ۲ الی شکل ۴ نمایش داده شده است. تحلیل این ماتریس‌ها در طبقه‌بندهای مختلف نشان می‌دهد که طبقه‌بند KNN در هر سه حالت طبقه‌بندی، کمترین مقدار FP را دارد، که بیانگر دقت بالاتر آن در کاهش هشدارهای کاذب است. پس از آن، SVM دارای مقدار FP بیشتری نسبت به KNN اما کمتر از GentleBoost است، در حالی که GentleBoost بیشترین مقدار FP را نشان می‌دهد. از سوی دیگر، مقدار FN در SVM کمترین مقدار در بین سه روش است، اما تفاوت آن با KNN بسیار ناچیز بوده و فقط یک نمونه

جدول ۳- مقایسه عملکرد طبقه‌بندها بر روی ویژگی‌های انتخاب‌شده در حالت‌های مختلف.

AUC	نمره F1	دقت	اختصاصیت	حساسیت	صحت		
۹۹/۰±۵۷/۲۳	۹۸/۰±۳۷/۵۳	۹۷/۰±۹۵/۸۰	۹۷/۰±۷۰/۹۲	۹۸/۰±۷۹/۶۵	۹۸/۰±۲۸/۵۶	SVM (Polynomial)	CTRL × PDOFF
۹۹/۰±۹۴/۰۶	۹۹/۰±۱۲/۳۱	۹۹/۰±۰۴/۳۴	۹۸/۰±۹۳/۳۸	۹۹/۰±۲۰/۵۲	۹۹/۰±۰۷/۳۳	SVM (RBF)	
۹۹/۰±۸۰/۲۲	۹۹/۰±۳۵/۲۴	۹۹/۰±۳۰/۳۴	۹۹/۰±۲۳/۳۹	۹۹/۰±۴۰/۴۳	۹۹/۰±۲۲/۲۵	KNN	
۸۸/۱±۶۵/۳۵	۸۷/۱±۶۲/۴۱	۸۷/۱±۶۸/۹۰	۸۶/۲±۳۲/۳۷	۸۷/۱±۵۹/۸۲	۸۶/۱±۹۸/۵۱	DT	
۹۹/۰±۵۳/۱۷	۹۶/۰±۸۰/۷۶	۹۷/۰±۳۵/۵۹	۹۷/۰±۰۹/۶۶	۹۶/۱±۲۶/۳۴	۹۶/۰±۶۵/۷۸	RF	
۹۹/۰±۶۲/۱۵	۹۷/۰±۵۲/۶۵	۹۷/۱±۳۵/۰۵	۹۷/۱±۰۳/۲۱	۹۷/۱±۷۱/۰۷	۹۷/۰±۳۹/۶۹	GentleBoost	
۹۸/۰±۳۶/۲۲	۹۳/۰±۷۲/۹۰	۹۲/۱±۲۰/۰۷	۹۱/۱±۰۵/۳۱	۹۵/۱±۳۰/۲۶	۹۳/۰±۲۸/۹۶	QDA	
۹۷/۳±۴۲/۵۱	۹۳/۴±۴۵/۷۷	۹۲/۴±۹۹/۷۱	۹۲/۵±۱۴/۲۹	۹۳/۴±۹۲/۹۳	۹۳/۵±۰۸/۰۴	MLP	
۲۱/۰±۵۶/۹۹	۳۲/۰±۳۲/۹۸	۴۹/۰±۹۲/۹۷	۵۶/۰±۶۸/۹۷	۴۷/۰±۷۲/۹۸	۳۴/۰±۲۲/۹۸	SVM (Polynomial)	CTRL × PDON
۰۶/۰±۹۲/۹۹	۳۱/۰±۹۹/۹۸	۴۱/۰±۸۲/۹۸	۴۶/۰±۶۹/۹۸	۵۴/۰±۱۵/۹۹	۳۲/۰±۹۳/۹۸	SVM (RBF)	
۱۹/۰±۷۳/۹۹	۲۴/۰±۰۹/۹۹	۳۲/۰±۶۶/۹۸	۳۶/۰±۵۰/۹۸	۳۲/۰±۵۲/۹۹	۲۶/۰±۰۴/۹۹	KNN	
۲۸/۱±۱۷/۸۹	۸۰/۰±۶۸/۸۸	۰۸/۱±۱۰/۸۹	۴۰/۱±۰۳/۸۸	۵۶/۱±۲۹/۸۸	۷۸/۰±۱۶/۸۸	DT	
۱۳/۰±۴۶/۹۹	۶۸/۰±۰۰/۹۷	۵۹/۰±۰۷/۹۷	۶۶/۰±۷۷/۹۶	۱۲/۱±۹۳/۹۶	۷۱/۰±۸۵/۹۶	RF	
۱۰/۰±۷۰/۹۹	۶۸/۰±۸۳/۹۷	۸۳/۰±۶۵/۹۷	۹۴/۰±۳۸/۹۷	۸۶/۰±۰۲/۹۸	۷۲/۰±۷۲/۹۷	GentleBoost	
۶۷/۰±۰۹/۹۸	۹۵/۰±۴۶/۹۳	۱۸/۱±۰۳/۹۲	۴۷/۱±۸۹/۹۰	۴۶/۱±۹۵/۹۴	۰۱/۱±۰۲/۹۳	QDA	
۸۳/۲±۲۴/۹۷	۴۴/۴±۷۸/۹۲	۲۲/۵±۲۳/۹۲	۱۶/۶±۳۱/۹۱	۷۸/۳±۲۶/۹۳	۸۰/۴±۳۴/۹۲	MLP	
۲۹/۰±۲۷/۹۹	۵۴/۰±۲۳/۹۷	۶۴/۰±۳۹/۹۷	۶۶/۰±۴۰/۹۷	۹۹/۰±۰۸/۹۷	۵۳/۰±۲۴/۹۷	SVM (Polynomial)	PDOFF × PDON
۰۸/۰±۹۰/۹۹	۴۳/۰±۶۵/۹۸	۶۸/۰±۵۶/۹۸	۶۹/۰±۵۵/۹۸	۶۶/۰±۷۴/۹۸	۴۳/۰±۶۵/۹۸	SVM (RBF)	
۱۵/۰±۶۵/۹۹	۳۶/۰±۴۹/۹۸	۵۹/۰±۴۸/۹۸	۶۰/۰±۴۸/۹۸	۵۸/۰±۵۰/۹۸	۳۶/۰±۴۹/۹۸	KNN	
۴۶/۱±۷۱/۸۴	۳۵/۱±۰۷/۸۳	۷۴/۱±۳۷/۸۳	۱۸/۲±۴۶/۸۳	۵۱/۲±۸۳/۸۲	۲۶/۱±۱۴/۸۳	DT	
۱۵/۰±۲۹/۹۹	۸۶/۰±۷۳/۹۵	۰۷/۱±۶۹/۹۵	۰۹/۱±۶۸/۹۵	۰۶/۱±۷۷/۹۵	۸۶/۰±۷۳/۹۵	RF	
۱۰/۰±۳۸/۹۹	۶۲/۰±۹۱/۹۵	۸۸/۰±۱۹/۹۶	۹۲/۰±۲۱/۹۶	۹۱/۰±۶۳/۹۵	۶۲/۰±۹۲/۹۵	GentleBoost	
۳۶/۰±۵۲/۹۶	۹۱/۰±۴۵/۸۹	۳۲/۱±۵۸/۹۲	۳۸/۱±۰۶/۹۳	۴۹/۱±۵۵/۸۶	۸۷/۰±۸۰/۸۹	QDA	
۲۹/۷±۱۰/۹۴	۹۳/۷±۸۵/۸۷	۷۳/۷±۱۵/۸۸	۴۴/۷±۲۸/۸۸	۲۸/۸±۵۸/۸۷	۷۶/۷±۹۳/۸۷	MLP	

جدول ۴: مقایسه عملکرد روش‌های تحلیل غیرخطی شامل آنتروپی نمونه‌ای (SampEn)، آنتروپی تقریبی (ApEn) و تحلیل کمی بازگشت.

PDON × PDOFF	CTRL × PDON	CTRL × PDOFF	
۹۸/۶۵±۰/۴۳	۹۹/۰۴±۰/۲۶	۹۹/۲۲±۰/۲۵	RQA
۵۴/۰±۵۳/۹۸	۳۵/۰±۹۴/۹۸	۳۲/۰±۹۰/۹۸	ApEn
۵۵/۰±۷۶/۹۸	۲۶/۰±۱۸/۹۹	۳۱/۰±۲۲/۹۹	SampEn

منحنی‌های ROC برای نمایش بصری و مقایسه عملکرد طبقه‌بندهای مختلف در سه حالت طبقه‌بندی استفاده شدند. این مطالعه منحنی‌های ROC را برای طبقه‌بندهای مختلف در سه حالت طبقه‌بندی دوتایی (CTRL × CTRL، CTRL × PDOFF، CTRL × PDON)

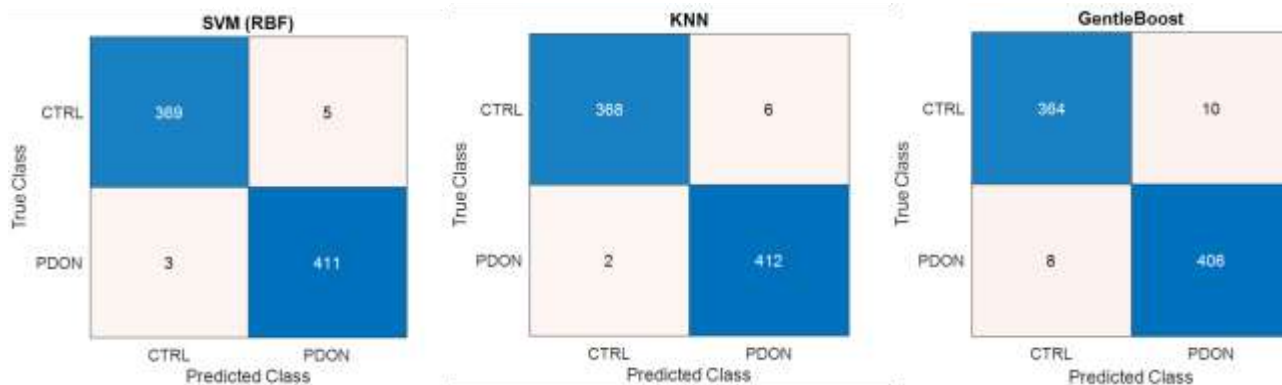
است. در حالی که SVM عملکردی نزدیک به KNN دارد، RF به‌طور عمومی در تمام معیارهای ارزیابی عملکرد ضعیف‌تری داشته است. نتایج نشان می‌دهند که KNN به‌عنوان یک مدل پیشرفته و با دقت بالا می‌تواند ابزار موثری برای تشخیص بیماری پارکینسون بر اساس داده‌های EEG باشد و همچنین نشان‌دهنده برتری الگوریتم‌های مبتنی بر KNN در مسائل طبقه‌بندی داده‌های پزشکی است.

به منظور ارائه جزئیات بیشتر از نتیجه طبقه‌بندی در طبقه‌بندی سه کلاسه، ماتریس‌های درهم‌ریختگی برای هر سه طبقه‌بند در شکل ۸ الی شکل ۱۰ نمایش داده شده است. در بین سه روش مذکور، طبقه‌بند KNN دارای بهترین عملکرد با بیشترین تشخیص درست را دارد. همچنین، طبقه‌بند SVM دارای بهترین عملکرد مناسبی است. در مجموع، طبقه‌بند RF دارای بدترین عملکرد در طبقه‌بندی سه کلاسه است.

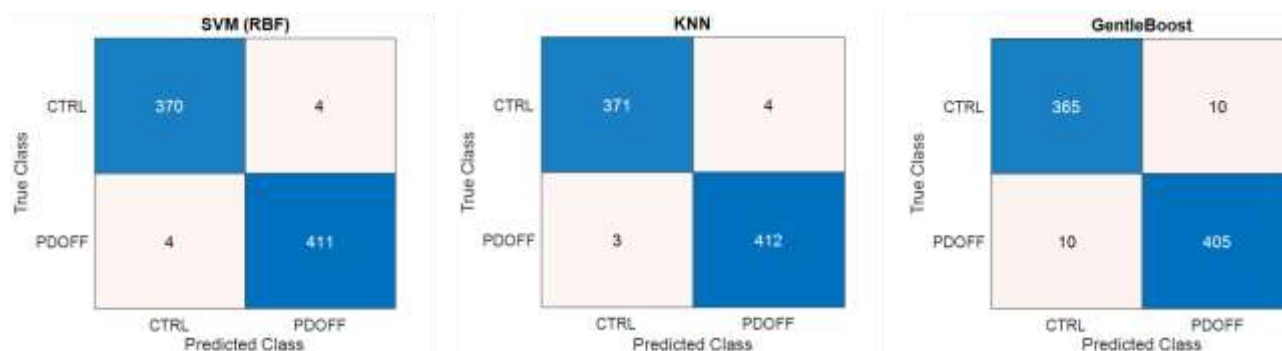
جدول ۶ صحت طبقه‌بندی روش پیشنهادی را با برخی از بهترین روش‌های ارائه شده در ادبیات، در سه حالت طبقه‌بندی (CTRL × PDOFF)، (CTRL × PDON) و (CTRL × PDON) مقایسه می‌کند. بر اساس نتایج جدول ۶، روش پیشنهادی در حالت‌های CTRL × PDOFF و CTRL × PDON با صحت‌های ۹۹/۳۲٪ و ۹۹/۰۴٪ نسبت به سایر روش‌های موجود در ادبیات عملکرد برتری نشان داده است. در

PDON و (PDOFF × PDON) در **Error! Reference source not found.** شکل ۷ ترسیم کرده است. همان‌طور که در شکل‌های مذکور نمایش داده شده است، طبقه‌بندهای SVM، KNN و GentleBoost عملکرد بهتری نسبت به سایر طبقه‌بندها از خود نشان دادند. با این حال، بر اساس معیار ROC، تفاوت معنادار آماری بین عملکرد طبقه‌بندهای SVM، KNN و RF مشاهده نشد.

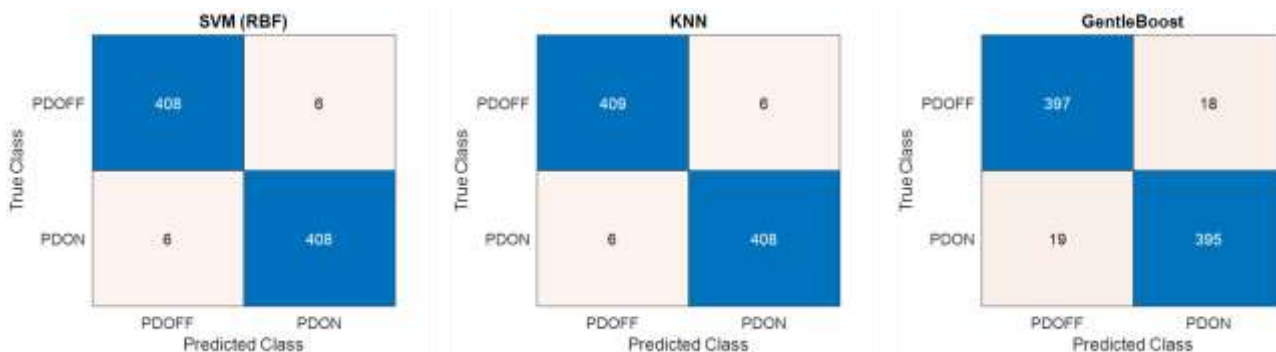
همچنین، در این مطالعه، عملکرد الگوریتم‌های طبقه‌بندی SVM، KNN و RF برای شناسایی بیماری پارکینسون بر اساس ویژگی‌های استخراج شده از سیگنال‌های EEG در حالت سه کلاسه مورد ارزیابی قرار گرفت. نتایج طبقه‌بندی به صورت خلاصه در جدول ۵ ارائه شده است. همان‌طور که مشاهده می‌شود، در طبقه‌بندی سه کلاسه الگوریتم KNN بهترین عملکرد را از نظر معیارهای مختلف ارزیابی از جمله صحت، دقت پیش‌بینی، حساسیت و اختصاصیت دارد. به‌طور خاص، میانگین کلان صحت KNN به میزان ۹۸/۳۱٪ است، که نسبت به الگوریتم SVM با میانگین صحت ۹۷/۷۰٪ و RF با میانگین صحت ۹۶/۲۸٪ بالاتر است. همچنین، طبقه‌بند KNN به‌طور قابل توجهی توانسته است حساسیت و اختصاصیت بالاتری را فراهم کند که نشان‌دهنده توانایی بهتر این مدل در شناسایی بیماران مبتلا به پارکینسون و تشخیص درست نمونه‌های سالم



شکل ۲- ماتریس‌های درهم‌ریختگی برای هر دسته‌بند در طبقه‌بندی دودویی CTRL × PDON.



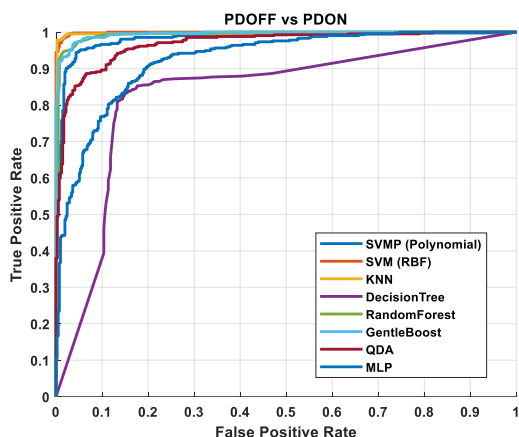
شکل ۳- ماتریس‌های درهم‌ریختگی برای هر دسته‌بند در طبقه‌بندی دودویی CTRL × PDOFF.



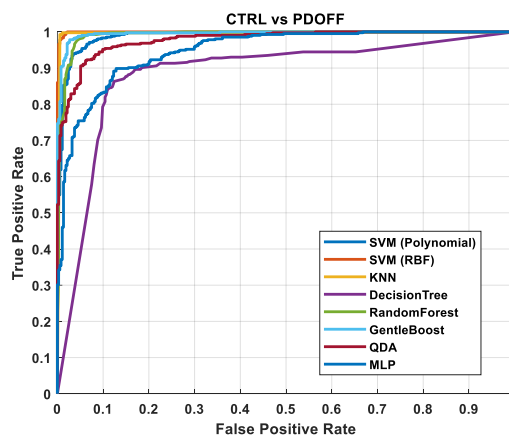
شکل ۴- ماتریس‌های درهم‌ریختگی برای هر دسته‌بند در طبقه‌بندی دودویی PDOFF × PDON.

جدول ۵- مقایسه عملکرد طبقه‌بندها بر روی ویژگی‌های انتخاب‌شده در طبقه‌بندی سه کلاسه شامل کلاس‌های CTRL، PDOFF و PDON.

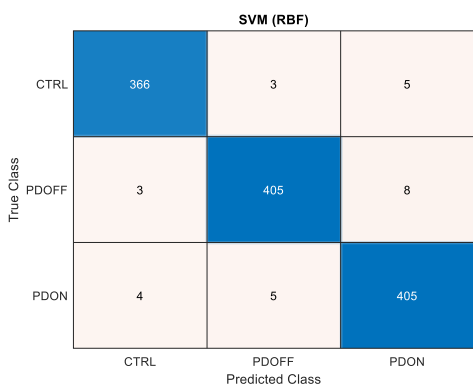
صحت	حساسیت	اختصاصیت	دقت	نمره F <sub>1</sub>		
۹۷/۷۰	۹۷/۷۰	۹۸/۸۴	۹۷/۷۲	۹۷/۷۱	میانگین کلان	SVM
۹۷/۷۱	۹۷/۷۰	۹۸/۸۵	۹۷/۷۰	۹۷/۷۰	میانگین خرد	
۹۸/۳۱	۹۸/۳۱	۹۹/۱۵	۹۸/۳۳	۹۸/۳۳	میانگین کلان	KNN
۹۸/۳۰	۹۸/۳۰	۹۹/۱۵	۹۸/۳۰	۹۸/۳۰	میانگین خرد	
۹۶/۲۸	۹۶/۲۸	۹۸/۱۴	۹۶/۲۷	۹۶/۲۷	میانگین کلان	RF
۹۶/۲۷	۹۶/۲۷	۹۸/۱۳	۹۶/۲۷	۹۶/۲۷	میانگین خرد	



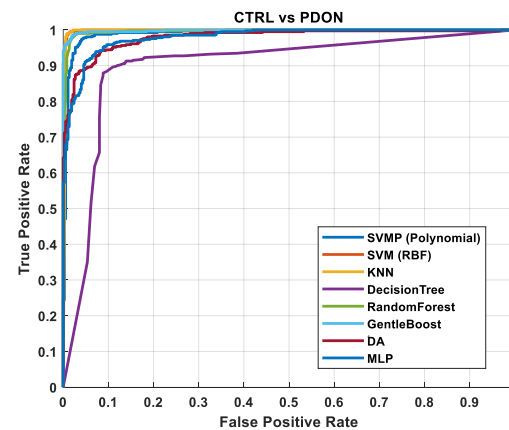
شکل ۷: منحنی‌های ROC برای هر دسته‌بند در طبقه‌بندی دودویی: PDOFF × PDON.



شکل ۵: منحنی‌های ROC برای هر دسته‌بند در طبقه‌بندی دودویی: CTRL × PDOFF.



شکل ۸: ماتریس درهم‌ریختگی برای طبقه‌بند SVM در دسته‌بندی سه کلاسه.



شکل ۶: منحنی‌های ROC برای هر دسته‌بند در طبقه‌بندی دودویی: CTRL × PDON.

هدف این پژوهش دسته‌بندی گروه کنترل (CTRL) در کنار گروه‌های بیماران (PDOFF و PDON) بود. بررسی ارتباط بین نواحی مختلف مغز ممکن است الگوهای ارزشمندی را برای شناسایی افراد مبتلا به بیماری پارکینسون آشکار کند. بنابراین، در تحقیقات آینده تمرکز بر تحلیل معیارهای ارتباطی بین نواحی مختلف مغز خواهد بود تا هم عملکرد طبقه‌بندی بهبود یابد و هم بینش عمیق‌تری درباره مکانیسم‌های عصبی زیربنایی بیماری پارکینسون ارائه شود.

#### ۴- نتیجه‌گیری

این مقاله یک رویکرد نوآورانه مبتنی بر یادگیری ماشینی برای تشخیص خودکار بیماری پارکینسون با استفاده از سیگنال‌های EEG ارائه کرده است. با استفاده از ویژگی‌های تحلیل کمی‌سازی بازگشتی استخراج‌شده از کانال‌های مختلف EEG که پیش‌پردازش شده‌اند، این روش به‌طور عمیق به بازیابی ویژگی‌های دینامیکی سیگنال‌ها می‌پردازد که به الگوهای متمایز فعالیت مغزی در مقایسه بین بیمار مبتلا به پارکینسون و فرد سالم اشاره دارند. انتخاب ویژگی مبتنی بر ReliefF موفق شد مرتبط‌ترین ویژگی‌ها را شناسایی کند. ویژگی‌های انتخاب‌شده همراه با آموزش طبقه‌بندی مختلف برای سه کاربرد طبقه‌بندی دوتایی و یک حالت طبقه‌بندی سه کلاسه بررسی شدند. در میان این طبقه‌بندی‌ها، KNN بهترین عملکرد را نشان داد و به میانگین صحت ۹۹/۳۲٪ در طبقه‌بندی دوتایی و میانگین صحت ۹۸/۳۲٪ در طبقه‌بندی سه کلاسه دست یافت، که نشان‌دهنده قابلیت اطمینان آن به‌عنوان یک ابزار تشخیصی است. این نتایج به‌وضوح کارایی روش پیشنهادی در استفاده از سیگنال‌های EEG برای تشخیص بیماری پارکینسون را نشان می‌دهد. این مطالعه سهم قابل توجهی در حوزه در حال ظهور تشخیص‌های غیرتهاجمی دارد و مسیرهای آینده‌ای را برای توسعه روش‌های تشخیصی زود هنگام و در پی آن رویکردهای درمانی بهتر پیشنهاد می‌دهد.

#### ۵- مراجع

[1] Titova, N., Qamar, M. A., & Chaudhuri, K. R. (2017). Biomarkers of Parkinson's disease: an introduction. *International Review of Neurobiology*, 132, 183-196.

[۲] نجفی آشتیانی محمد، طاهری ماهرخ. (۲۰۲۴). تحلیل روایی و پایایی دستگاه طراحی و ساخته شده برای سنجش

True Class	Predicted Class		
	CTRL	PDOFF	PDON
CTRL	368	2	4
PDOFF	2	407	5
PDON	2	5	407

شکل ۹: ماتریس درهم‌ریختگی برای طبقه‌بند KNN در دسته‌بندی سه کلاسه.

True Class	Predicted Class		
	CTRL	PDOFF	PDON
CTRL	361	6	7
PDOFF	9	396	10
PDON	7	6	401

شکل ۱۰: ماتریس درهم‌ریختگی برای طبقه‌بند RF در دسته‌بندی سه کلاسه.

جدول ۶- مقایسه صحت طبقه‌بندی بین روش پیشنهادی و برخی از بهترین روش‌های ارائه شده در ادبیات.

PDON × PDOFF	CTRL × PDON	CTRL × PDOFF	
۶۸/۶۷	۸۷/۷۸	۸۹/۵۶	مقاله [۷]
۹۹/۲۰	-	-	مقاله [۲۳]
-	-	۸۵/۴۰	مقاله [۲۴]
۹۸/۹۷	۹۸/۸۵	۹۹/۰۱	مقاله [۲۵]
۹۸/۷۹	۹۹/۲۰	۹۸/۶۲	مقاله [۲۶]
۹۸/۴۹	۹۹/۰۴	۹۹/۳۲	روش پیشنهادی

حالت PDON × PDOFF نیز با صحت ۹۸/۴۹٪ عملکرد رقابتی و قابل قبولی نسبت به سایر روش‌ها ارائه شده است. این نتایج بیانگر کارایی بالای روش پیشنهادی در طبقه‌بندی دقیق گروه‌های مورد بررسی و برتری نسبی آن در مقایسه با روش‌های پیشین می‌باشد.



computational tools (PRED+ CT). *Frontiers in neuroinformatics*, 11, 67.

[16] Furman, Ł., Duch, W., Minati, L., & Tołpa, K. (2023). Short-time Fourier transform and embedding method for recurrence quantification analysis of EEG time series. *The European Physical Journal Special Topics*, 232(8), 1513-1527.

[17] Lau, Z. J., Pham, T., Chen, S. A., & Makowski, D. (2022). Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations. *European Journal of Neuroscience*, 56(7), 5047-5069.

[18] Acharya, U. R., Sree, S. V., Chattopadhyay, S., Yu, W., & Ang, P. C. A. (2011). Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. *International journal of neural systems*, 21(03), 199-211.

[19] Heunis, T., Aldrich, C., Peters, J. M., Jeste, S. S., Sahin, M., Scheffer, C., & De Vries, P. J. (2018). Recurrence quantification analysis of resting state EEG signals in autism spectrum disorder—a systematic methodological exploration of technical and demographic confounders in the search for biomarkers. *BMC medicine*, 16, 1-17.

[20] Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6), 237-329.

[21] Chen, Y., & Pham, T. D. (2013). Sample entropy and regularity dimension in complexity analysis of cortical surface structure in early Alzheimer's disease and aging. *Journal of neuroscience methods*, 215(2), 210-217.

[22] Urbanowicz, R. J., Olson, R. S., Schmitt, P., Meeker, M., & Moore, J. H. (2018). Benchmarking relief-based feature selection methods for bioinformatics data mining. *Journal of biomedical informatics*, 85, 168-188.

[23] Shah, S. A. A., Zhang, L., & Bais, A. (2020). Dynamical system based compact deep hybrid network for classification of Parkinson disease related EEG signals. *Neural Networks*, 130, 75-84.

[24] Anjum, M. F., Dasgupta, S., Mudumbai, R., Singh, A., Cavanagh, J. F., & Narayanan, N. S. (2020). Linear predictive coding distinguishes spectral EEG features of Parkinson's disease. *Parkinsonism & related disorders*, 79, 79-85.

[25] Aljalal, M., Aldosari, S. A., AlSharabi, K., Abdurraqeab, A. M., & Alturki, F. A. (2022). Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques. *Diagnostics*, 12(5), 1033.

[26] Aljalal, M., Aldosari, S. A., Molinas, M., AlSharabi, K., & Alturki, F. A. (2022). Detection of Parkinson's disease from EEG signals using discrete wavelet transform, different entropy measures, and machine learning techniques. *Scientific Reports*, 12(1), 22547.

دقت حس عمقی مج پا در افراد مبتلا به پارکینسون. *نشریه‌ی علمی مهندسی پزشکی زیستی*, ۱۱(۱)، ۲۰-۲۰.

[۳] حسنی، & مالکی. (۲۰۲۲). اثر پیش‌رفت بیماری پارکینسون بر ویژگی‌های آشوبی RQA. *نشریه‌ی علمی مهندسی پزشکی زیستی*, ۱۶(۳)، ۲۰۷-۲۱۸.

[4] Ma, Z. L., Wang, Z. L., Zhang, F. Y., Liu, H. X., Mao, L. H., & Yuan, L. (2024). Biomarkers of Parkinson's disease: from basic research to clinical practice. *Aging and Disease*, 15(4), 1813.

[5] Altham, C., Zhang, H., & Pereira, E. (2024). Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's Disease: A systematic review. *Plos one*, 19(5), e0303644.

[6] Maitín, A. M., García-Tejedor, A. J., & Muñoz, J. P. R. (2020). Machine learning approaches for detecting Parkinson's disease from EEG analysis: a systematic review. *Applied Sciences*, 10(23), 8662.

[7] Coelho, B. F. O., Massaranduba, A. B. R., dos Santos Souza, C. A., Viana, G. G., Brys, I., & Ramos, R. P. (2023). Parkinson's disease effective biomarkers based on Hjorth features improved by machine learning. *Expert Systems with Applications*, 212, 118772.

[8] Dhivyaa, C. R., Nithya, K., & Anbukkarasi, S. (2024). Enhancing Parkinson's Disease Detection and Diagnosis: A Survey of Integrative Approaches Across Diverse Modalities. *IEEE Access*.

[9] Liao, X. Y., Gao, Y. X., Qian, T. T., Zhou, L. H., & Li, L. Q. (2024). Bibliometric analysis of electroencephalogram research in Parkinson's disease from 2004 to 2023. *Frontiers in Neuroscience*, 18, 1433583.

[10] Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2018). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, 1-7.

[11] Latifoğlu, F., Penekli, S., Orhanbulucu, F., & Chowdhury, M. E. (2024). A novel approach for Parkinson's disease detection using Vold-Kalman order filtering and machine learning algorithms. *Neural Computing and Applications*, 36(16), 9297-9311.

[12] Koch, M., Geraedts, V., Wang, H., Tannemaat, M., & Bäck, T. (2019, December). Automated machine learning for EEG-based classification of Parkinson's disease patients. In *2019 IEEE international conference on big data (Big Data)* (pp. 4845-4852). IEEE.

[13] Alawee, W. H., Basem, A., & Al-Haddad, L. A. (2023). Advancing biomedical engineering: Leveraging Hjorth features for electroencephalography signal analysis. *Journal of Electrical Bioimpedance*, 14(1), 66-72.

[14] De Oliveira, A. P. S., De Santana, M. A., Andrade, M. K. S., Gomes, J. C., Rodrigues, M. C., & dos Santos, W. P. (2020). Early diagnosis of Parkinson's disease using EEG, machine learning and partial directed coherence. *Research on Biomedical Engineering*, 36, 311-331.

[15] Cavanagh, J. F., Napolitano, A., Wu, C., & Mueen, A. (2017). The patient repository for EEG data+