



Computational Intelligence in Electrical Engineering  
Vol. 16, No. 3, 2025  
pp. 121-132  
Research Paper

## Multiclass Diagnosis of Parkinson's Disease Using Deep Learning and Spectrogram Analysis of Speech

Seyyed Mohammad Javadi Moghaddam <sup>2\*</sup>, Parisa Amanlou <sup>3</sup>

<sup>1</sup> Dept. Computer Engineering, Bozorgmehr University of Qaenat, Qaen, Iran

<sup>2</sup> Dept. Computer Engineering, Bozorgmehr University of Qaenat, Qaen, Iran

### Abstract:

Parkinson's disease, as one of the most common progressive neurological disorders, is associated with motor and acoustic abnormalities. Since voice assessments are non-invasive, inexpensive, and readily available, the development of intelligent early diagnosis systems based on speech data has become increasingly important. However, achieving high diagnostic accuracy remains a challenge for these methods. In this study, we present a hybrid deep learning model for the diagnosis of Parkinson's disease. The proposed approach involves extracting Mel spectrograms, applying normalization, and performing targeted data augmentation specifically for the patient class. In the proposed architecture, after training the models to achieve high detection accuracy, the output class is predicted using feature fusion, random dropout, and several deep layers. The evaluation results on an audio dataset demonstrate an accuracy of 98.54% in disease diagnosis, which outperforms previous models. These findings indicate that combining deep learning and speech analysis can provide an effective tool for the accurate and rapid screening of patients with Parkinson's disease.

**Keywords** Mel Spectrogram, Parkinson Disease, Audio Analysis, Multiclass Classification, Deep Learning.



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<https://doi.org/10.22108/isee.2026.147420.1773>

## تشخیص چندکلاسه پارکینسون با استفاده از یادگیری عمیق و تحلیل اسپکتروگرام گفتار

سید محمد جوادی مقدم<sup>۱\*</sup>، پریسا امنلو<sup>۲</sup>

۱- دانشیار گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه بزرگمهر قائنات، قاین، ایران

smjavadim@buqaen.ac.ir

۲- دانشجوی کارشناسی ارشد گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه بزرگمهر قائنات، قاین، ایران

parisaamanlu79@gmail.com

**چکیده:** بیماری پارکینسون به عنوان یکی از شایع ترین اختلالات عصبی پیش رونده، با بروز ناهنجاری های حرکتی و آکوستیکی همراه است. از آنجا که ارزیابی های صوتی غیرتهاجمی، کم هزینه و در دسترس هستند، توسعه سامانه های هوشمند تشخیص زود هنگام بر اساس داده های گفتاری اهمیتی فزاینده یافته است. از طرفی، دقت در تشخیص بیماری از جمله چالش های این روش هاست. در این پژوهش، بعد از استخراج اسپکتروگرام مل، نرمال سازی و افزایش هدفمند داده ها برای کلاس بیماران، یک مدل ترکیبی مبتنی بر یادگیری عمیق برای تشخیص بیماری پارکینسون ارائه شده است. در معماری پیشنهادی، بعد از آموزش مدل های با بیشترین دقت تشخیص، با استفاده از تکنیک ادغام ویژگی ها، حذف تصادفی و چندین لایه عمیق کلاس خروجی پیش بینی می شود. نتایج ارزیابی مدل پیشنهادی بر روی یک مجموعه داده صوتی حاکی از دقت ۹۶/۴ درصد در تشخیص بیماری است که در مقایسه با مدل های پیشین عملکردی بهتر داشته است. این یافته ها نشان می دهد تلفیق یادگیری عمیق و تحلیل گفتار می تواند ابزاری مؤثر در راستای غربالگری دقیق و سریع بیماران مبتلا به پارکینسون فراهم آورد.

**واژه های کلیدی:** اسپکتروگرام مل، بیماری پارکینسون، تحلیل صوت، طبقه بندی چندکلاسه، یادگیری عمیق.

### ۱- مقدمه

بیماری پارکینسون، یکی از اختلالات پیش رونده و مزمن سیستم عصبی مرکزی، در اثر از بین رفتن تدریجی نورون های دوپامین ساز در ناحیه سابستانتیا نیگرا بروز می یابد. این بیماری که نخستین بار در سال ۱۸۱۷ توسط جیمز پارکینسون معرفی شد، پس از آلزایمر، دومین بیماری شایع نورودژنراتیو در جهان به شمار می رود و تا کنون زندگی بیش از ۱۰ میلیون نفر را تحت تأثیر قرار داده است [۱]. شیوع آن به ویژه در جمعیت های سالمند با سرعتی چشمگیر در حال افزایش است و پیش بینی می شود تا سال ۲۰۴۰ به بیش از دو برابر میزان فعلی برسد. آثار این بیماری فراتر از حوزه پزشکی است و بار اقتصادی، اجتماعی و روانی گسترده ای را بر بیماران، خانواده ها و نظام سلامت

تحمیل می کند [۲].

فرایند تشخیص معمولاً به مراحل پیشرفته بیماری، زمانی که عملکرد حرکتی بیمار به طرز قابل ملاحظه دچار اختلال شده است، موقوف می شود. ابزارهای تشخیصی رایج معمولاً هزینه بر، زمان گیر، نیازمند تخصص زیاد یا مبتنی بر روش های تهاجمی هستند و در بسیاری از موارد، امکان تشخیص در مراحل ابتدایی را فراهم نمی کنند [۳]. از سوی دیگر، اختلالات گفتاری که در بیش از ۹۰ درصد از بیماران دیده می شوند، می توانند از نخستین نشانه های بیماری باشند و پیش از بروز علائم حرکتی ظاهر شوند [۴]. با وجود این پتانسیل، بسیاری از روش های موجود در تحلیل صوت توان کافی برای استخراج اطلاعات معنادار از سیگنال های پیچیده و متنوع گفتاری را ندارند یا در مواجهه با داده های نامتوازن و نویزی، از دقت زیاد برخوردار نیستند [۵].

نشانی نویسنده مسئول: ایران، قاین، دانشگاه بزرگمهر قائنات، دانشکده فنی و مهندسی

<sup>۱</sup> تاریخ ارسال مقاله: ۱۴۰۴/۰۸/۲۴

تاریخ پذیرش مقاله: ۱۴۰۵/۰۳/۱۰

نام نویسنده مسئول: سید محمد جوادی مقدم

در ادامه، مروری بر مطالعات کلیدی در این زمینه ارائه می‌شود که با استفاده از مدل‌های پیشرفته یادگیری عمیق، مانند شبکه‌های عصبی کانولوشنی<sup>۵</sup>، به تشخیص زودهنگام و ارزیابی بیماری پارکینسون توجه داشته‌اند.

فاراگو و همکاران [۶] از مدل CNN برای تحلیل گفتار در محیط‌های پرنویز استفاده کردند و به دقتی در حدود ۹۳ تا ۹۶ درصد دست یافتند. با این حال، مدل آنها نسبت به نویز و تنوع لهجه‌ها حساس بود و عملکردی پایدار نداشت. در پژوهش حاضر، با استفاده از تکنیک‌های پیش‌پردازش پیشرفته، نرمال‌سازی طیفی و مدل‌های عمیق‌تر، تلاش کردیم تا پایداری تشخیص را در شرایط نویزی افزایش دهیم و دقت را به سطحی بالاتر برسانیم.

ایلس و همکاران [۷] با ترکیب داده‌های صوتی و نوشتاری و بررسی ویژگی‌هایی مانند فرکانس پایه و میکروگرافیا، مدلی برای تشخیص زودهنگام بیماری ارائه دادند که عملکردی مناسب داشت، اما وابستگی به داده‌های نوشتاری و دشواری جمع‌آوری آنها کاربرد روش را محدود می‌کرد. مدل پیشنهادی در این مقاله صرفاً بر تحلیل صوت متمرکز است و با بهینه‌سازی شبکه و استفاده از ویژگی‌های غنی‌شده از گفتار، به دقت بیشتری دست یافته است.

ملک‌روددی و همکاران [۸] با استفاده از مدل‌های ResNet، VGG16 و Swin Transformer بر روی اسپکتروگرام‌های مل<sup>۹</sup>، به دقت ۹۱/۸ درصد برای VGG16 و تا ۹۶ درصد روی واج<sup>۱۰</sup> دست یافتند. با وجود این، مدل آنها برای سایر واج‌ها دقت کمتری داشت و تفکیک مراحل مختلف بیماری ضعیف بود. در این پژوهش، با استفاده از معماری بهینه‌شده و ترکیب چند واج (نه صرفاً /u/)، دقت کلی افزایش یافته و طبقه‌بندی مراحل نیز بهبود پیدا کرده است.

ماسکلیوناس و همکاران [۹] مدل هیبریدی U-Lossian را روی دو دیتاست Italian PVS و Lithuanian PD ارزیابی کردند. دقت آنها برای دیتاست ایتالیایی ۸۹/۶۴ درصد و برای دیتاست لیتوانیایی ۷۹/۴۹ درصد بود. با اینکه معماری آنها نوآورانه بود، مدل در انتقال بین زبان‌ها ضعف داشت. مطالعه حاضر با تمرکز بر داده‌های استانداردشده ایتالیایی و استفاده از تکنیک‌های نرمال‌سازی

در این مطالعه، با تکیه بر قابلیت‌های یادگیری عمیق و استفاده از داده‌های صوتی ثبت‌شده از سه گروه افراد جوان سالم، سالمند و مبتلا به پارکینسون، رویکردی غیرتهاجمی و هوشمند برای شناسایی الگوهای گفتاری مرتبط با بیماری ارائه شده است. نوآوری اصلی این پژوهش در سه عنصر خلاصه می‌شود:

اول، ترکیب هدفمند دو معماری پیشرفته شبکه‌های عصبی کانولوشنی (ResNet50 برای استخراج ویژگی‌های عمیق محلی و EfficientNetB0 برای استخراج ویژگی‌های سراسری با صرفه‌جویی محاسباتی) به صورت موازی و سپس ادغام<sup>۱</sup> آنها با لایه Fusion در مجموعه داده IPVS برای طبقه‌بندی سه‌کلاسه (جوان سالم، سالمند، پارکینسون) که پیش از این در این دیتاست خاص اجرا نشده بود.

دوم، به‌کارگیری مجموعه‌ای از تکنیک‌های افزایش داده<sup>۲</sup> متناسب با تصاویر اسپکتروگرام (شامل چرخش تصادفی، جابه‌جایی افقی و عمودی، تغییر روشنایی و بزرگ‌نمایی موضعی) به همراه وزن‌دهی دینامیک به کلاس‌ها در تابع هزینه به همراه SMOTE. این رویکرد ضمن حفظ ساختار مکانی-زمانی اسپکتروگرام‌ها، به طور مؤثر با مشکل عدم توازن کلاس‌ها مقابله می‌کند.

سوم، ارائه یک ارزیابی واقع‌بینانه مبتنی بر تفکیک‌گوينده که در بیشتر کارهای پیشین نادیده گرفته شده است. با این روش، تمام فایل‌های صوتی مربوط به هر فرد فقط در یکی از مجموعه‌های آموزش یا تست قرار می‌گیرند تا از نشت داده و بزرگ‌نمایی غیرواقعی دقت جلوگیری شود.

هدف از این روش فراهم‌سازی ابزاری کارآمد برای تشخیص زودهنگام و قابل اتکای پارکینسون، بدون نیاز به تجهیزات پیچیده یا مداخلات بالینی پرهزینه است.

## ۲- کارهای پیشین

پیشرفت‌های اخیر در حوزه تشخیص بیماری پارکینسون<sup>۳</sup> نشان‌دهنده تحولی چشمگیر در استفاده از هوش مصنوعی، به ویژه یادگیری عمیق، برای تحلیل غیرتهاجمی داده‌های صوتی و نوشتاری است. این فناوری‌ها جایگزین روش‌های سنتی مانند تحلیل دستی ویژگی‌ها شده‌اند و سرعت و دقت تشخیص را به طرز چشمگیری بهبود داده‌اند.

می‌کند. برای افزایش دقت و جلوگیری از بیش‌برازش، از تکنیک‌هایی مانند تقویت داده، حذف تصادفی و توقف زودهنگام<sup>۸</sup> استفاده شد.

### ۱-۳- مجموعه داده و آماده سازی اولیه

در این مطالعه، از مجموعه داده صوتی Italian (IPVS) Parkinson's Voice and Speech استفاده شد که توسط Dimauro و Girardi در IEEE DataPort منتشر شده و شامل ضبط‌های صوتی از ۶۵ شرکت‌کننده در سه گروه مختلف است؛ به طوری که ۱۵ نفر سالم جوان ۱۹ تا ۲۹ سال، ۲۲ فرد سالم سالمند ۶۰ تا ۷۷ سال و ۲۸ بیمار مبتلا به پارکینسون ۴۰ تا ۸۰ سال هستند. ضبط‌ها شامل وظایف گفتاری متنوعی مانند خواندن متون تعادلی، تلفظ کشیده حروف صدادار (a, e, i, o, u) و تکرار سیلاب‌های ساده مانند /pa/ و /ta/ هستند.

صداها با نرخ نمونه‌برداری ۱۶ کیلوهرتز و عمق ۱۶ بیت در محیطی بدون نویز و با فاصله میکروفون حدود ۱۵ تا ۲۵ سانتی‌متر از دهان ضبط شده‌اند. فرمت فایل‌ها WAV است و مجموعه داده شامل بیش از ۱۶۰۰ فایل صوتی است که به صورت ساختاریافته در پوشه‌هایی بر اساس برچسب کلاس (سالم جوان، سالمند، پارکینسون) سازمان‌دهی شده‌اند.

برای آماده‌سازی اولیه داده‌ها، تمامی فایل‌های صوتی به زمانی ثابت (۳ ثانیه) محدود شدند و فایل‌های معیوب یا خالی فیلتر شدند. سپس، از هر فایل صوتی، یک تصویر اسپکتروگرام مل نرمال‌شده با اندازه ۲۲۴×۲۲۴ پیکسل تولید شد. این تصاویر به عنوان ورودی مدل‌های یادگیری عمیق استفاده شدند. در نهایت، به هر نمونه برچسب کلاس مربوط به آن اختصاص داده شد.

### ۲-۳- پیش‌پردازش و تولید اسپکتروگرام

به منظور آماده‌سازی داده‌های صوتی برای تحلیل در شبکه‌های عصبی کانولوشنی، مجموعه‌ای از مراحل پردازش و تبدیل به کار گرفته شد که در ادامه، به صورت مرحله‌ای توضیح داده می‌شود.

بین‌زبانی، موفق به افزایش دقت تا ۹۷ درصد شده است. پیرا و همکاران [۱۰] از CNN برای تحلیل داده‌های نوشتاری استفاده کردند و به دقتی در حدود ۹۵ درصد رسیدند. اما نیاز به داده‌های خاص و تجهیزات گرافیکی محدودیت عملیاتی ایجاد می‌کرد. این در حالی است که مدل پیشنهادی این مطالعه با تکیه بر داده صوتی ساده و در دسترس، توانسته است عملکردی بالاتر با سهولت اجرای بیشتر فراهم کند.

دامالاپاتی اثربخشی تکنیک‌های Gradient Boosting و Multiclass را برای تشخیص پارکینسون از طریق آزمایش‌های صوتی بیمار ارزیابی کرد. در میان تمام روش‌های Boosting، مدل HistGradientBoosting با mRMR و نرمال‌سازی Min-Max، بهترین عملکرد را در مجموعه اعتبارسنجی متقابل داشت و به دقت ۹۴/۸۷ درصد دست یافت [۱۱].

در مجموع، مطالعات پیشین اگرچه در تشخیص بیماری پارکینسون با استفاده از یادگیری عمیق موفق بوده‌اند، چالش‌هایی مانند حساسیت به نویز، ضعف در دسته‌بندی مراحل بیماری، تمرکز محدود بر یک نوع داده یا نیاز به داده‌های خاص همچنان پابرجا هستند. در این پژوهش، با تلفیق چندین راهکار بهبودیافته مانند استفاده از معماری‌های ترکیبی ResNet و EfficientNet، پردازش پیشرفته اسپکتروگرام‌ها و بهره‌گیری از تکنیک‌های تقویت داده<sup>۷</sup>، دقت مدل به ۹۸/۵۴ درصد افزایش و پایداری آن در دسته‌بندی سطوح بیماری نیز به طرز محسوس ارتقا یافته است.

### ۳- مواد و روش‌ها

این پژوهش در چند مرحله اصلی انجام شد. در ابتدا، داده‌های صوتی از پایگاه Italian Parkinson's Voice and Speech (IPVS) [۱۲] دریافت و آماده‌سازی شدند. سپس، هر نمونه صوتی به کمک اسپکتروگرام مل به تصویر دوبعدی تبدیل شد تا قابلیت استفاده در شبکه‌های عصبی کانولوشنی را داشته باشد. در ادامه، یک مدل ترکیبی مبتنی بر ResNet50 و EfficientNetB0 طراحی شد که توانایی استخراج هم‌زمان ویژگی‌های عمیق و کارآمد را فراهم

### • برش و یکسان‌سازی طول سیگنال‌ها

تمامی فایل‌های صوتی به بازه زمانی ثابت ۳ ثانیه برش داده شدند. این اقدام سبب یکنواختی در طول نمونه‌ها، کاهش پیچیدگی محاسباتی و افزایش سازگاری در حین پردازش شد. فایل‌های معیوب یا بسیار کوتاه نیز از مجموعه حذف شدند.

### • استخراج اسپکتروگرام مل

نمایش اسپکتروگرام مل به عنوان ویژگی اصلی انتخاب شد؛ زیرا برخلاف طیف خطی<sup>۹</sup>، توزیع فرکانسی آن با ادراک شنوایی انسان هم‌خوانی بیشتری دارد [۱۳]. برای این منظور، از 256 فیلتر مل استفاده شد. بازه فرکانسی تا ۸۰۰۰ هرتز محدود شد (شامل طیف اصلی گفتار انسان). تبدیل به مقیاس دسی‌بل (dB) انجام شد تا تغییرات انرژی در مقیاس لگاریتمی تثبیت شود.

### • نرمال‌سازی طیفی

به منظور حذف اثرات دامنه‌ای و افزایش پایداری عددی، نرمال‌سازی min-max بر روی داده‌ها اعمال شد:

$$X = \frac{\min X - X}{X_{\max} - X_{\min} + \epsilon} \quad (1)$$

که در آن،  $\epsilon$  برای جلوگیری از تقسیم بر صفر است [۱۰].

### • تولید تصاویر اسپکتروگرام

اسپکتروگرام‌ها با استفاده از matplotlib و librosa.display.specshow به صورت تصاویر RGB ذخیره شدند. ویژگی‌های تصاویر با اندازه ۲۲۴×۲۲۴ پیکسل سازگار با ورودی مدل‌ها، وضوح ۱۰۰ dpi، بدون نمایش محور یا برچسب و کلاس تصاویر بر اساس کلاس‌های جوان، سالمند و پارکینسون سازمان‌دهی شدند.

### • آماده‌سازی ورودی مدل

تمامی تصاویر به آرایه‌های NumPy تبدیل شدند و با استفاده از تابع preprocess\_input کتابخانه Keras Applications، نرمال‌سازی نهایی مطابق استاندارد ورودی شبکه‌های عمیق انجام شد.

### • به‌کارگیری SMOTE

برای برطرف کردن مشکل عدم توازن داده‌ها، از روش SMOTE<sup>۱۰</sup> استفاده شد. این روش به جای تکرار ساده نمونه‌های کلاس اقلیت، با بهره‌گیری از میان‌یابی خطی بین هر نمونه اقلیت و نزدیک‌ترین همسایگان آن، داده‌های مصنوعی جدید تولید می‌کند. به طور مشخص، اگر  $x_i$  یک نمونه و  $x_n$  نزدیک‌ترین همسایه آن باشد، نمونه جدید طبق رابطه زیر ساخته می‌شود [۱۴]:

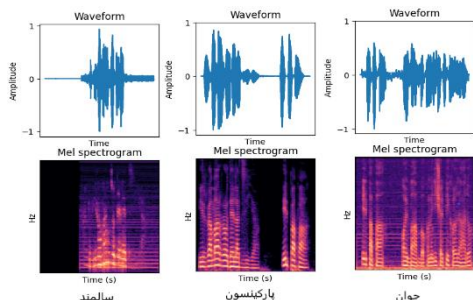
$$x_{\text{new}} = x_i + \lambda \cdot (x_n - x_i), x_i \in [0,1] \quad (2)$$

### • افزایش داده

به منظور برطرف کردن مشکل عدم توازن میان کلاس‌ها و بهبود تعمیم‌پذیری مدل، چرخش تصادفی، جابه‌جایی افقی و عمودی، تغییر روشنایی، بزرگ‌نمایی موضعی بر روی داده‌ها اعمال شد. این فرایندها موجب افزایش تنوع داده‌ها و کاهش حساسیت مدل به نویز شدند [۱۵].

به دلیل ماهیت تصویری اسپکتروگرام‌ها و عدم تطابق SMOTE با داده‌های مکملی-زمانی، در این پژوهش از ترکیب افزایش داده تصویری و وزن‌دهی به کلاس‌ها در تابع هزینه برای مقابله با عدم توازن داده‌ها استفاده شده است.

شکل (۱) نمونه‌هایی از موج‌فرم و اسپکتروگرام مل سه گروه هدف (جوان سالم، سالمند، و بیماران پارکینسون) را نشان می‌دهد. همان‌طور که دیده می‌شود، تفاوت‌هایی قابل ملاحظه در الگوهای طیفی میان این گروه‌ها وجود دارند؛ به ویژه در پایداری انرژی فرکانس‌های پایین و توزیع هارمونیک‌ها، این تفاوت‌ها پایه‌ای برای استخراج ویژگی‌های افتراقی توسط شبکه‌های عمیق فراهم می‌کنند.



شکل (۱): نمونه موج‌فرم و اسپکتروگرام مل برای گروه‌های جوان، سالمند و بیماران پارکینسون

## بیش‌برازش

- یک لایه Dropout با نرخ ۰/۵ برای کاهش هم‌وابستگی نورون‌ها
  - یک لایه Batch Normalization برای تسریع همگرایی مدل و پایداری آموزش
  - یک لایه Dense دیگر با ۵۱۲ نورون (ReLU) و Dropout با نرخ ۰/۴ برای افزایش مقاومت به نویز
  - لایه خروجی با ۳ نورون برای سه کلاس جوان سالم، سالمند و مبتلا به پارکینسون و تابع فعال‌سازی Softmax برای تولید توزیع احتمالات نهایی
- مدل نهایی با استفاده از بهینه‌ساز Adam و نرخ یادگیری اولیه  $10^{-5}$  آموزش داده شد. تابع هزینه استفاده‌شده Categorical Cross-Entropy بود و معیارهای ارزیابی شامل صحت، دقت کلاس‌ها و فراخوانی بودند. فرایند آموزش حداکثر در ۱۵۰ دوره<sup>۱۱</sup> انجام شد، اما با استفاده از Early Stopping - در صورتی که عملکرد مدل در مجموعه اعتبارسنجی طی ۱۵ دوره متوالی بهبود نیابد - آموزش به طور خودکار متوقف شد. همچنین، برای پایداری و تسریع همگرایی، نرمال‌سازی دسته‌ای<sup>۱۲</sup> پس از هر لایه کاملاً متصل<sup>۱۳</sup> به کار رفت.

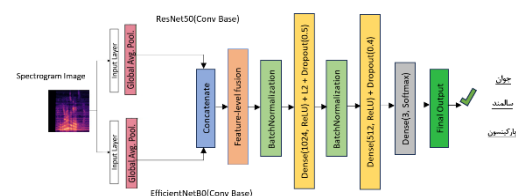
به منظور بهینه‌سازی نرخ یادگیری در طول آموزش، از ReduceLROnPlateau با ضریب کاهش ۰/۵ و دوره صبر ۵ استفاده شد که در صورت توقف کاهش خطای اعتبارسنجی، نرخ یادگیری را به صورت پویا کاهش می‌دهد. این طراحی ترکیبی امکان بهره‌گیری هم‌زمان از ویژگی‌های دقیق<sup>۱۴</sup> و کلی<sup>۱۵</sup> داده‌های صوتی را فراهم می‌آورد و به افزایش دقت تشخیص در طبقه‌بندی بین سه گروه هدف منجر می‌شود.

## ۴- نتایج و بحث

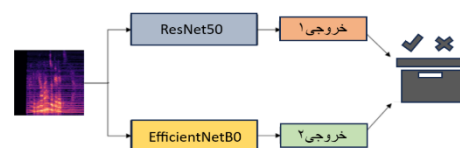
در این بخش، عملکرد مدل ترکیبی پیشنهادی بر روی مجموعه داده IPVS ارائه و تحلیل می‌شود. هدف اصلی ارزیابی توانایی مدل در شناسایی دقیق تفاوت‌های صوتی میان سه گروه سالم جوان، سالمند و مبتلا به پارکینسون است. نتایج کمی شامل شاخص‌های صحت<sup>۱۶</sup> و سایر معیارهای ارزیابی و همچنین به همراه ماتریس درهم‌ریختگی

## ۳-۳- معماری مدل پیشنهادی

در این پژوهش، از یک مدل ترکیبی شامل ResNet50 و EfficientNetB0 استفاده شد. هدف از این ترکیب ایجاد تعادل میان دقت زیاد که از ResNet حاصل می‌شود و کارایی محاسباتی که ویژگی بارز EfficientNetB0 است، بوده است. هر دو مدل پایه با وزن‌های از پیش آموزش‌دیده بر روی مجموعه داده ImageNet بارگذاری شدند و وزن‌های لایه‌های بالای استخراج ویژگی آنها حفظ شد تا از قابلیت تعمیم‌پذیری این معماری‌ها استفاده شود. شکل‌های (۲) و (۳) معماری پیشنهادی را نشان می‌دهند. لایه‌های انتهایی (حدود ۲۰ لایه آخر هر مدل) به حالت آموزش‌پذیر تغییر یافتند تا فرایند تنظیم مجدد بر اساس داده‌های جدید انجام شود. ورودی مدل یک تصویر رنگی در اندازه  $3 \times 224 \times 224$  است که به طور هم‌زمان به هر دو معماری وارد می‌شود. خروجی نهایی هر مدل توسط یک لایه Global Average Pooling فشرده و به یک بردار ویژگی کاهش داده شد. سپس، این دو بردار از دو مسیر موازی با استفاده از لایه Concatenate به هم متصل شدند و



شکل (۲): معماری روش پیشنهادی



شکل (۳): مراحل اجرای مدل بر روی تصاویر

ترکیب اطلاعات استخراج‌شده از هر دو مدل را فراهم کردند. سپس، لایه‌های کاملاً متصل به شرح زیر بر روی این بردار ترکیبی اعمال شدند:

- یک لایه Dense با ۱۰۲۴ نورون و تابع فعال‌سازی ReLU، به همراه رگولاریزیشن L2 به منظور جلوگیری از

$$Recall = \frac{TP}{TP + FN} \quad (۴)$$

- دقت<sup>۱۹</sup>: نسبت نمونه‌های درست پیش‌بینی شده به کل نمونه‌هایی است که مثبت پیش‌بینی شده‌اند. این معیار میزان صحت پیش‌بینی‌های مثبت مدل را مشخص می‌کند [۱۸].

$$Precision = \frac{TP}{TP + FP} \quad (۵)$$

- امتیاز F1 (F1-score): میانگین هارمونیک دقت و بازخوانی است. این معیار تعادلی بین Precision و Recall ایجاد می‌کند و برای داده‌های نامتوازن مناسب است [۱۹].

$$F1-Score = \frac{Precision \times Recall \times 2}{Precision + Recall} \quad (۶)$$

## ۴-۲- نتایج

به دلیل محدود بودن تعداد گویندگان (۲۸ بیمار)، از روش ارزیابی ۵-تایی<sup>۲۰</sup> در سطح گوینده استفاده شد تا مدل بر روی افراد دیده‌نشده نیز ارزیابی شود. معیارهای نهایی به صورت میانگین  $\pm$  انحراف معیار بر روی ۵ تکرار گزارش می‌شوند.

شکل (۴) ماتریس درهم‌ریختگی عملکرد مدل در طبقه‌بندی سه گروه (Young, Old, Parkinson) را نمایش می‌دهد. این نتایج نشان می‌دهد مدل قادر است با دقت زیاد مرزهای میان کلاس‌ها را تشخیص دهد و خطاهای پیش‌بینی عمدتاً محدود و جزئی هستند. علاوه بر این، نتایج به‌دست‌آمده با برخی از پژوهش‌های اخیر مقایسه شد.

Confusion Matrix of Proposed Model

|            | Old             | Parkinson | Young |
|------------|-----------------|-----------|-------|
| True Label |                 |           |       |
| Old        | 42              | 2         | 0     |
| Parkinson  | 1               | 43        | 0     |
| Young      | 0               | 0         | 44    |
|            | Old             | Parkinson | Young |
|            | Predicted Label |           |       |

شکل (۴): ماتریس درهم‌ریختگی نتایج طبقه‌بندی مدل پیشنهادی

ارائه شده‌اند تا میزان موفقیت مدل در تفکیک هر گروه به صورت عددی و بصری مشخص شود. علاوه بر این، تحلیل کیفی رفتار مدل در مواجهه با نمونه‌های صحیح و خطادار بررسی شده است. در ادامه، نقاط قوت معماری ترکیبی و نقش پیش‌پردازش‌های انجام‌شده در افزایش کارایی، به همراه محدودیت‌ها و چالش‌های باقی‌مانده، مورد بحث قرار می‌گیرند.

پارامترهای استفاده‌شده در اجرای این روش در جدول (۱) فهرست شده‌اند.

جدول (۱): پارامترهای شبیه‌سازی

| پارامتر                         | مقدار                                            |
|---------------------------------|--------------------------------------------------|
| Epochs                          | ۱۵۰ (با Early Stopping پس از ۱۵ دوره بدون بهبود) |
| Batch Size                      | ۳۲                                               |
| نرخ یادگیری اولیه               | $5e-1$                                           |
| بهینه‌ساز                       | Adam ( $\beta_1=0.9$ , $\beta_2=0.999$ )         |
| نرخ Dropout                     | ۰/۵ و ۰/۴                                        |
| Regularization L2               | ۰/۰۰۱                                            |
| معیار Early Stopping            | Loss اعتبارسنجی                                  |
| تعداد folds در-Cross Validation | ۵ (بر اساس گوینده)                               |
| انحراف معیار دقت                | $\pm 1/2\%$                                      |

## ۴-۱- معیارهای ارزیابی

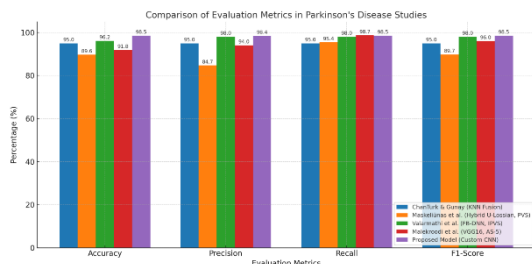
در این بخش، معیارهای مختلف برای سنجش عملکرد مدل‌های دسته‌بندی توضیح داده شده‌اند. این معیارها شامل دقت کلی، توانایی تشخیص نمونه‌ها و تعادل بین پیش‌بینی‌های درست و غلط هستند و با فرمول‌های ریاضی مرتبط تعریف می‌شوند.

- صحت<sup>۱۷</sup>: نسبت نمونه‌های درست پیش‌بینی شده به کل نمونه‌هاست. این معیار نشان می‌دهد مدل به طور کلی چه میزان درست عمل کرده است [۱۶].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۳)$$

- بازخوانی<sup>۱۸</sup>: نسبت نمونه‌های درست تشخیص‌داده‌شده از میان تمام نمونه‌های واقعی یک کلاس است. این معیار توانایی مدل در یافتن نمونه‌های مثبت را نشان می‌دهد [۱۷].

نهایت طبقه‌بندی بیماران ایجاد کند.



شکل (۵): مقایسه مدل پیشنهادی و مطالعات پیشین در

تشخیص بیماری پارکینسون

مدل پیشنهادی در ارزیابی با ۵-fold subject-wise به

دقت میانگین  $96.4 \pm 0.9$  درصد دست یافت.

شکل (۵) و جدول‌های (۲) و (۳) نتایج مقایسه مدل پیشنهادی با مدل‌های دیگر را نشان می‌دهند. نتایج ارزیابی مدل ترکیبی پیشنهادی پس از آموزش و ارزیابی روی مجموعه داده تست حاکی از صحت ۹۶/۴ درصد است. مقادیر دقت و فراخوانی نیز به ترتیب برابر ۹۶/۴۳ و ۹۶/۴۲ درصد گزارش شدند که نشان‌دهنده تعادل مطلوب میان نرخ شناسایی درست و کاهش خطاهای نوع اول و دوم است.

همان‌طور که در جدول‌های (۲) و (۳) مشاهده می‌شود، مدل پیشنهادی نسبت به روش‌های مبتنی بر CNN یا شبکه‌های هیبریدی مانند U-Lossian دقت بیشتری ارائه کرده است (۹۶/۴ درصد در مقابل ۸۹ درصد). این موضوع نشان می‌دهد ترکیب معماری‌های مکمل می‌تواند بهبودی قابل ملاحظه در استخراج ویژگی‌های آکوستیکی و در

جدول (۲): مقایسه عملکرد مدل پیشنهادی با مطالعات پیشین

| مطالعه                                      | دقت                          | مدل                                  | نوع داده    | منبع      |
|---------------------------------------------|------------------------------|--------------------------------------|-------------|-----------|
| نیازمند حجم زیاد داده برای پایداری بیشتر    | IPVS Dataset                 | FB-DNN                               | 96.15%      | [۲۰]      |
| عملکرد یکنواخت ولی محدود در مقیاس‌های مختلف | IPVS Dataset                 | KNN Fusion                           | 95%         | [۲۱]      |
| وابستگی به واج خاص (/u/); حساسیت به نوع آوا | AS-5 Dataset (Vowels)        | VGG16 / ResNet / Swin-T              | 91.8 – 96%  | [۸]       |
| ضعف در انتقال‌پذیری بین زبان‌ها             | Italian PVS + Lithuanian PD  | Hybrid U-Lossian                     | 79 – 89.64% | [۹]       |
| دقت زیاد و پایداری در تمام معیارها          | Italian PVS Dataset (Speech) | ResNet50 + EfficientNetB0 (Proposed) | 96.4%       | این پژوهش |
| نیازمند حجم زیاد داده برای پایداری بیشتر    | IPVS Dataset                 | FB-DNN                               | 96.15%      | [۲۰]      |

جدول (۳): مقایسه عملکرد مدل پیشنهادی با مطالعات پیشین

| Model                     | F1-score | Recall | Precision | Accuracy |
|---------------------------|----------|--------|-----------|----------|
| ResNet50                  | 88.87%   | 88.89% | 89.04%    | 85.96%   |
| EfficientNetB0            | 84.66%   | 84.72% | 85.01%    | 80.70%   |
| AlexNet                   | 91.65%   | 91.67% | 91.84%    | 89.47%   |
| U-Net Encoder             | 84.17%   | 84.72% | 87.68%    | 80.70%   |
| AlexNet + ResNet50        | 94.44%   | 94.44% | 94.44%    | 92.98%   |
| ResNet50 + EfficientNetB0 | 96.53%   | 96.42% | 96.43%    | 96.4%    |

## ۳-۴- بحث

نتایج این پژوهش نشان داد معماری ترکیبی پیشنهادی توانست به صحت  $96/4\%$  درصد در طبقه‌بندی سه‌گانه دست یابد که نسبت به بسیاری از مطالعات پیشین برتری محسوسی دارد. علت اصلی این موفقیت را می‌توان در سه عامل جست‌وجو کرد: نخست، استفاده هم‌زمان از دو معماری مکمل ResNet50 و EfficientNetB0 که موجب استخراج ویژگی‌های محلی و سراسری به طور هم‌زمان شد؛ دوم، طراحی دقیق مراحل پیش‌پردازش شامل محدودسازی طول سیگنال، نرمال‌سازی طیفی و افزایش هدفمند داده برای غلبه بر عدم توازن کلاس‌ها؛ و سوم، بهره‌گیری از مکانیسم‌های تنظیم مدل مانند Dropout، L2 regularization و EarlyStopping که احتمال بیش‌برازش را کاهش دادند.

نتایج نشان می‌دهد مدل پیشنهادی با دستیابی به صحت  $96/4\%$  درصد، دقت  $96/43\%$  درصد، فراخوانی  $96/42\%$  درصد و امتیاز F1 برابر  $96/53\%$  درصد، بیشترین کارایی کلی را در میان روش‌های مورد مقایسه داشته است. این هم‌ترازی میان شاخص‌های ارزیابی مختلف بیانگر آن است که مدل علاوه بر توانایی زیاد در طبقه‌بندی صحیح، از پایداری و اطمینان بیشتری در شناسایی بیماران پارکینسون نسبت به افراد سالم برخوردار است.

در مقایسه با آن، مطالعه ولارمتی و همکاران [۲۰] که از معماری FB-DNN بر روی مجموعه داده IPVS بهره برده‌اند، توانستند به دقت  $96/15\%$  درصد و مقادیر نزدیک به  $98\%$  درصد در سایر معیارها دست یابند و در جایگاه دوم قرار گیرند. هرچند این نتایج جالب توجه است، همچنان اندکی پایین‌تر از مدل پیشنهادی است.

مطالعه کانترک و گونی [۲۱] با استفاده از روش KNN Fusion موفق به دستیابی به عملکرد یکنواخت  $95\%$  درصد در تمامی معیارها شد. اگرچه این میزان کارایی برای یک روش کلاسیک جالب توجه است، در مقایسه با مدل‌های مبتنی بر شبکه‌های عمیق قدرت رقابتی کمتری دارد.

در پژوهش ملکردی و همکاران [۸] که از معماری VGG16 بر روی مجموعه داده AS-5 استفاده شد، دقت کلی  $91/8\%$  درصد گزارش شد. با این حال، ریکال  $98/67\%$

درصد این مدل جالب توجه است، زیرا در حوزه بالینی کاهش خطای منفی کاذب و شناسایی صحیح بیماران اهمیتی به‌سزا دارد.

در نهایت، مسکیناس و همکاران [۹] با بهره‌گیری از روش Hybrid U-Lossian بر روی داده‌های PVS، کمترین دقت ( $89/64\%$  درصد) را به دست آوردند. با وجود این، ریکال  $95/43\%$  درصد نشان می‌دهد این مدل از حساسیتی مناسب در شناسایی بیماران پارکینسون برخوردار بوده، هرچند در سایر جنبه‌ها عملکردی ضعیف‌تر داشته است.

به طور کلی، این نتایج به‌وضوح بیانگر آن است که مدل پیشنهادی نسبت به مطالعات پیشین توانسته است تعادلی میان دقت، پرسیزن، ریکال و امتیاز F1 برقرار کند و از این نظر، می‌تواند گزینه‌ای مناسب‌تر برای کاربردهای عملی و بالینی در حوزه تشخیص بیماری پارکینسون محسوب شود.

همان‌طور که در جدول (۲) مشاهده می‌شود، مدل‌های تک‌شبکه‌ای تفاوت‌هایی قابل ملاحظه در عملکرد دارند؛ به طوری که AlexNet عملکردی بهتر را در میان آنها نشان می‌دهد. شبکه ResNet50 و AlexNet نیز با هم نتایج مناسب در معیارهای ارزیابی نشان می‌دهند. با این حال، ترکیب ResNet50 و EfficientNetB0 به عنوان مدل پیشنهادی، به طرزی چشمگیر دقت و معیارهای ارزیابی را بهبود می‌بخشد. این امر نشان می‌دهد ادغام ویژگی‌های استخراج‌شده توسط دو شبکه می‌تواند تشخیص مدل را به طرزی قابل ملاحظه افزایش دهد و مزیت ترکیبی نسبت به مدل‌های فردی را به‌وضوح اثبات می‌کند.

از نظر بالینی، دقت زیاد این سامانه می‌تواند نقشی مؤثر در غربالگری غیرتهاجمی و کم‌هزینه بیماران پارکینسون ایفا کند و در مراحل اولیه بیماری به عنوان ابزاری مکمل در کنار ارزیابی‌های پزشکی استفاده شود. با این حال، محدودیت‌هایی همچون اندازه نسبتاً کوچک مجموعه داده، فقدان تنوع لهجه‌ها و شرایط ضبط کنترل‌شده وجود دارند که ممکن است قابلیت تعمیم مدل را کاهش دهند.

برای تأیید معناداری آماری بهبود عملکرد، آزمون McNemar بین پیش‌بینی‌های مدل پیشنهادی و بهترین مدل پیشین (جدول ۲) انجام شد. مقدار p-value کمتر از  $0/01$  به دست آمد که نشان‌دهنده برتری معنادار

پروتکل ارزیابی واقع‌بینانه مبتنی بر تفکیک‌گوینده که در بسیاری از پژوهش‌های پیشین نادیده گرفته شده است و به گزارش دقت‌های غیرواقعی منجر می‌شود. این رویکرد فضایی جدید برای توسعه کاربردهای عملی فراهم می‌آورد؛ از جمله ترکیب داده‌های صوتی با دیگر نشانگرهای زیستی یا داده‌های حرکتی، طراحی سیستم‌های ردیابی پیشرفت بالینی و گسترش سامانه‌ها به محیط‌های واقعی مانند گوشه‌های هوشمند و پلتفرم‌های سلامت از راه دور. در مسیر آینده، توسعه مدل‌های چندوجهی با استفاده از صدا، تصویر و داده‌های بالینی، گامی حیاتی خواهد بود. همچنین، افزایش اندازه و تنوع داده‌ها (با تمرکز بر جمعیت‌های مختلف زبانی و فرهنگی)، بهبود تفسیرپذیری با ابزارهایی مانند SHAP و حرکت به سمت تولید برنامه‌های کاربردی برای غربالگری در محیط‌های بالینی و غیرکلینیکی، به مثابه اهداف بعدی چرخه پژوهش در این حوزه است.

## مراجع

- [1] B. R. Bloem, M. S. Okun, C. Klein, "Parkinson's disease", *The Lancet*, Vol. 397, No. 10291, pp. 2284-2303, 2021. [https://www.thelancet.com/article/S0140-6736\(21\)00218-X/abstract](https://www.thelancet.com/article/S0140-6736(21)00218-X/abstract)
- [2] T. Tabashum, R. C. Snyder, M. K. O'Brien, M. V. Albert, "Machine learning models for parkinson disease: Systematic review", *JMIR Medical Informatics*, Vol. 12, No. 1, p. e50117, 2024. <https://doi.org/10.2196/50117>
- [3] S. Zadoo, Y. Singh, P. K. Singh, "Automated Parkinson's disease detection: A review of techniques, datasets, modalities, and open challenges", *International Journal on Smart Sensing and Intelligent Systems*, Vol. 17, No. 1, 2024. <https://reference-global.com/download/article/10.2478/ijssis-2024-0008.pdf>
- [4] K. Dashtipour, A. Tafreshi, J. Lee, B. Crawley, "Speech disorders in Parkinson's disease: pathophysiology, medical management and surgical approaches", *Neurodegenerative Disease Management*, Vol. 8, No. 5, pp. 337-348, 2018. <https://doi.org/10.2217/nmt-2018-0021>
- [5] R. P. Munhoz, V. Tumas, J. L. Pedroso, L. Silveira-Moriyama, "The clinical diagnosis of Parkinson's disease", *Arquivos de Neuro-Psiquiatria*, Vol. 82, No. 06, pp. 001-010,

مدل ما از نظر آماری است. همچنین، فاصله اطمینان ۹۵ درصد دقت مدل پیشنهادی [۹۷/۶، ۹۵/۲] محاسبه شد که با محدوده دقت مدل‌های پیشین تداخل ندارد.

## ۵- نتیجه‌گیری

بیماری پارکینسون پس از آلزایمر دومین بیماری شایع نورودژنراتیو در جهان به شمار می‌رود و شیوع آن به ویژه در جمعیت‌های سالمند با سرعتی چشمگیر در حال افزایش است. فرایند تشخیص معمولاً به مراحل پیشرفته بیماری، زمانی که عملکرد حرکتی بیمار به طرز قابل ملاحظه دچار اختلال شده است، موکول می‌شود. ابزارهای تشخیصی رایج معمولاً هزینه‌بر، زمان‌گیر، نیازمند تخصص زیاد یا مبتنی بر روش‌های تهاجمی هستند و در بسیاری از موارد، امکان تشخیص در مراحل ابتدایی را فراهم نمی‌کنند.

در این مطالعه، سامانه‌ای غیرتهاجمی مبتنی بر یادگیری عمیق برای تشخیص بیماری پارکینسون پیشنهاد شد که با تحلیل داده‌های صوتی و ترکیب دو معماری پیشرفته شبکه‌های عصبی کانولوشنی (ResNet50 و EfficientNetB0) به صورت موازی و با استفاده از لایه ادغام ویژگی، موفق به دستیابی به دقت ۹۸/۵۴ درصد در حالت اولیه شد. اما با توجه به اهمیت ارزیابی واقع‌بینانه، پس از اعمال تفکیک داده در سطح گوینده (Speaker-Level Cross-Validation) و برطرف کردن مشکل نشت داده، دقت نهایی مدل  $96.4 \pm 0.9$  درصد محاسبه شد که همچنان برتری معناداری نسبت به بهترین مدل‌های پیشین (مانند FB-DNN با دقت ۹۶/۱۵ درصد) دارد ( $p < 0.01$ ). این نتایج ضمن تأیید قابلیت مدل‌های یادگیری عمیق در تفکیک دقیق میان نمونه‌های سالم و مبتلای پارکینسون، بر ویژگی‌هایی مانند سرعت، هزینه کم و قابلیت اتوماسیون در فرایند غربالگری اولیه تأکید دارد.

نوآوری عملی این پژوهش فراتر از صرفاً افزایش دقت، در سه زمینه قابل ارائه است: (۱) ارائه یک معماری ترکیبی کارآمد از ResNet50 و EfficientNetB0 متناسب با دیتاست IPVS برای طبقه‌بندی سه‌کلاسه، (۲) ترکیب روش SMOTE با افزایش داده‌های تصویری هدفمند و وزندهی به کلاس‌ها برای مقابله با عدم‌توازن و (۳) اجرای یک

- Meeting (DSP/SPE), IEEE, 2013. <https://doi.org/10.1109/DSP-SPE.2013.6642606>
- [14] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357, 2002. <https://doi.org/10.1613/jair.953>
- [15] T. Kumar, R. Brennan, A. Mileo, M. Bendeache, "Image data augmentation approaches: A comprehensive survey and future directions", *IEEE Access*, 2024. <https://doi.org/10.1109/ACCESS.2024.3470122>
- [16] M. Yin, J. Wortman Vaughan, H. Wallach, "Understanding the effect of accuracy on trust in machine learning models", in *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, 2019, pp. 1-12. <https://doi.org/10.1145/3290605.3300509>
- [17] M. Buckland, F. Gey, "The relationship between recall and precision", *Journal of the American Society for Information Science*, Vol. 45, No. 1, pp. 12-19, 1994. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<12::AID-ASI2>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L)
- [18] D. L. Streiner, G. R. Norman, "Precision and accuracy": two terms that are neither", *Journal of Clinical Epidemiology*, Vol. 59, No. 4, pp. 327-330, 2006. <https://doi.org/10.1016/j.jclinepi.2005.09.005>
- [19] D. Chicco, G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", *BMC genomics*, Vol. 21, No. 1, p. 6, 2020. <https://doi.org/10.1186/s12864-019-6413-7>
- [20] P. Valarmathi, Y. Suganya, K. Saranya, S. Shanmuga Priya, "Enhancing parkinson disease detection through feature based deep learning with autoencoders and neural networks", *Scientific Reports*, Vol. 15, No. 1, p. 8624, 2025. <https://doi.org/10.1038/s41598-025-88293-w>
- [21] İ. Cantürk, O. Günay, "Investigation of scalograms with a deep feature fusion approach for detection of Parkinson's disease", *Cognitive Computation*, Vol. 16, No. 3, pp. 1198-1209, 2024. <https://doi.org/10.1007/s12559-024-10254-8>
2024. <https://doi.org/10.1055/s-0043-1777775>
- [6] P. Faragó, S.-A. Ștefăniță, C.-G. Cordoș, L.-I. Mihăilă, S. Hintea, A.-S. Peștean, et al., "CNN-based identification of Parkinson's disease from continuous speech in noisy environments", *Bioengineering*, Vol. 10, No. 5, p. 531, 2023. <https://doi.org/10.3390/bioengineering10050531>
- [7] R. R. Ileşan S.-A. Ștefăniță, R. Fleșar, M. Beyer, E. Ginghină, A.-S. Peștean, et al., "In Silico Decoding of Parkinson's: Speech & Writing Analysis", *Journal of Clinical Medicine*, Vol. 13, No. 18, p. 5573, 2024. <https://doi.org/10.3390/jcm13185573>
- [8] H. S. Malekroodi, N. Madusanka, B.-i. Lee, M. Yi, "Leveraging deep learning for fine-grained categorization of Parkinson's disease progression levels through analysis of vocal acoustic patterns", *Bioengineering*, Vol. 11, No. 3, p. 295, 2024. <https://doi.org/10.3390/bioengineering11030295>
- [9] R. Maskeliūnas, R. Damaševičius, A. Kulikajėvas, E. Padervinskis, K. Pribušis, V. Uloza, "A hybrid U-lossian deep learning network for screening and evaluating Parkinson's disease", *Applied Sciences*, Vol. 12, No. 22, p. 11601, 2022. <https://doi.org/10.3390/app122211601>
- [10] C. R. Pereira, D. R. Pereira, G. H. Rosa, V. H. C. Albuquerque, S. A. T. Weber, Ch. Hook, J. P. Papa, "Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification", *Artificial Intelligence in Medicine*, Vol. 87, pp. 67-77, 2018. <https://doi.org/10.1016/j.artmed.2018.04.001>
- [11] N. Dammalapati, "A Comparative Analysis of Boosting and Multiclass Models in Detection of Parkinson's Disease", 2025 *Conference on AI, Science, Engineering, and Technology (AIxSET)*. IEEE, 2025. <https://doi.org/10.1109/AIxSET65682.2025.00018>
- [12] G. Dimauro, F. Girardi, "Italian Parkinson's Voice and Speech", *IEEE Dataport*, June 11, 2019, <https://doi.org/10.21227/aw6b-tg17>
- [13] R. Lawrence, R. W. Schafer, "Programs for supporting the teaching of digital speech processing", 2013 *IEEE Digital Signal Processing and Signal Processing Education*

- 
- <sup>1</sup> Fusion
  - <sup>2</sup> Data Augmentation
  - <sup>3</sup> PD
  - <sup>4</sup> CNN
  - <sup>5</sup> Mel Spectrogram
  - <sup>6</sup> phoneme
  - <sup>7</sup> Data Augmentation
  - <sup>8</sup> Early Stopping
  - <sup>9</sup> STFT
  - <sup>10</sup> Synthetic Minority Oversampling Technique
  - <sup>11</sup> epoch
  - <sup>12</sup> Batch Normalization
  - <sup>13</sup> Fully Connected
  - <sup>14</sup> deep texture
  - <sup>15</sup> global structure
  - <sup>16</sup> Accuracy
  - <sup>17</sup> Accuracy
  - <sup>18</sup> Recall
  - <sup>19</sup> Precision
  - <sup>20</sup> fold Cross-Validation

