خلاصه ای از مقاله های منبع

فهرست

[مقاله ی Islam et al., 2025 3](#_Toc208518327)

[مدل 3](#_Toc208518328)

[ویژگی ها 7](#_Toc208518329)

[داده 8](#_Toc208518330)

[دیتاست 8](#_Toc208518331)

[آماده سازی داده 8](#_Toc208518332)

[نتایج 10](#_Toc208518333)

[مقایسه با مدل های دیگر 11](#_Toc208518334)

[بحث 11](#_Toc208518335)

[مقاله ی Rahmatallah et al., 2025 12](#_Toc208518336)

[مدل 13](#_Toc208518337)

[ویژگی ها 13](#_Toc208518338)

[دیتاست‌ها 14](#_Toc208518339)

[نتیجه و مقایسه با مطالعات پیشین 15](#_Toc208518340)

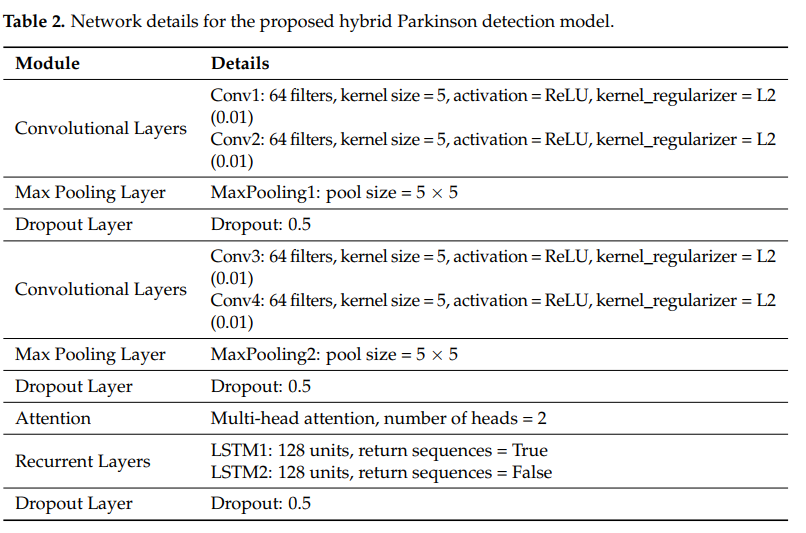
# مقاله ی Islam et al., 2025

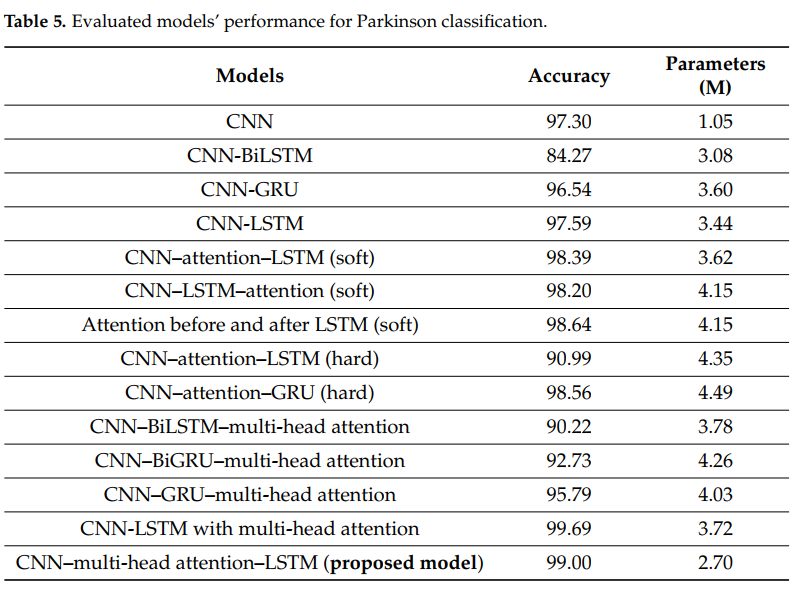
در مطالعه‌ی Islam و همکاران (2025) تحت عنوان “PD-Net: Parkinson’s Disease Detection Through Fusion of Two Spectral Features Using Attention-Based Hybrid Deep Neural Network”، نویسندگان با بهره‌گیری از داده‌های صوتی بیماران پارکینسون و افراد سالم، رویکردی نوین برای تشخیص این بیماری ارائه کرده‌اند. این مقاله که در ژورنال Information منتشر شده، با ترکیب ویژگی‌های طیفی Mel Spectrogram و MFCC و طراحی یک مدل هیبریدی عمیق، به دقت 99 درصد در طبقه‌بندی روی دیتاست Italina Parkinson’s Voice and Speech دست یافته است.

## مدل

رویکرد پیشنهادی مقاله، شبکه‌های عصبی کانولوشنی (**CNN**) و یک **multi-head attention mechanism** را همراه با لایه‌های **LSTM** ترکیب می‌کند تا همبستگی‌های فضایی و زمانی موجود در سیگنال صوتی شرکت‌کنندگان شناسایی شود. در ابتدا، داده‌های ورودی از طریق لایه‌های **CNN** پردازش می‌شوند تا ویژگی‌های فضایی استخراج و الگوهای محلی شناسایی شوند. سپس یک لایه **pooling** برای کاهش ابعاد و در عین حال حفظ اطلاعات ضروری به‌کار گرفته می‌شود تا بهره‌وری محاسباتی بهبود یابد.

خروجی تخت‌شده‌ی **CNN** به یک **multi-head attention mechanism** داده می‌شود که به مدل امکان می‌دهد بر بخش‌های مهم‌تر دنباله تمرکز کند و روابط پیچیده را از طریق attention heads متنوع شناسایی کند. در ادامه، داده‌ی پردازش‌شده توسط attention به لایه‌های **LSTM** با ۱۲۸ واحد داده می‌شود تا وابستگی‌های زمانی بلندمدت در سیگنال گفتار ثبت گردد. در پایان، پیش از رسیدن به لایه‌ی خروجی، خروجی تخت‌شده‌ی CNN، خروجی attention و خروجی LSTM در یک **feature vector** ادغام شده و سپس از طریق لایه‌های **dense** پردازش می‌شوند. فرایند **classification** بااستفاده از **sigmoid activation function** پیش‌بینی می‌کند که ورودی متعلق به یک فرد سالم (label = 0) یا بیمار پارکینسون (label = 1) است. برای افزایش **robustness** و جلوگیری از **overfitting**، تکنیک‌هایی مانند **dropout** و **L2 regularization** در طول آموزش به‌کار گرفته می‌شوند. ساختار دقیق مدل در جدول2 مقاله نشان داده شده است.

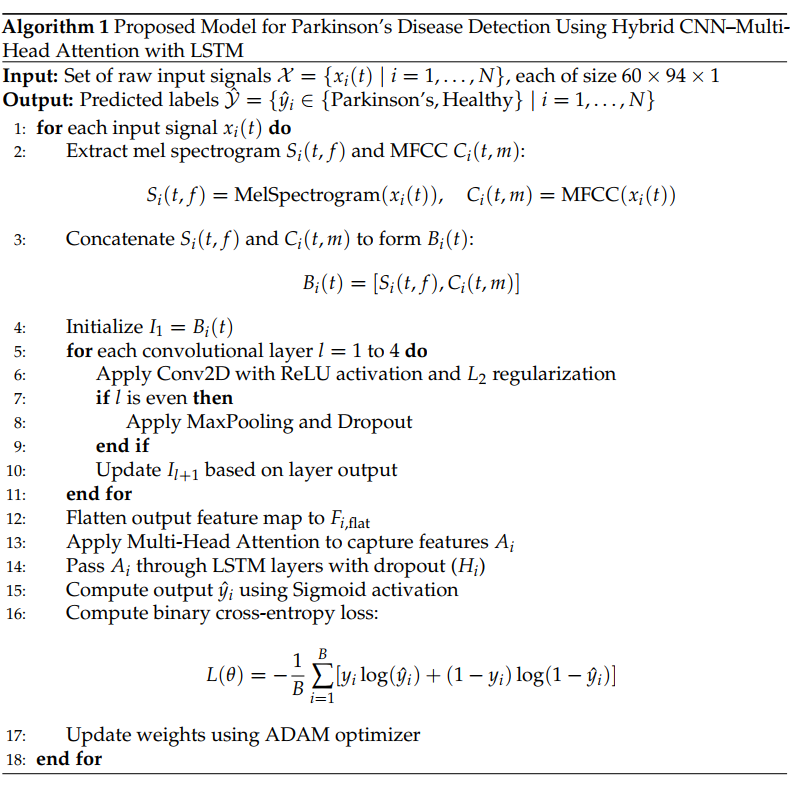


در بخش مقایسه مدل‌ها، عملکرد روش‌های مختلف از نظر دقت و تعداد پارامترها بررسی شده است. مدل پایه **CNN** با 1.05 میلیون پارامتر دقت 97.30٪ را به دست آورده است، در حالی که مدل‌های **CNN-BiLSTM** و **CNN-GRU** نتایج پایداری ارائه نکرده‌اند. افزودن **multi-head attention** باعث بهبود چشمگیر عملکرد شد؛ برای نمونه، مدل **CNN-LSTM + attention** با 3.72 میلیون پارامتر به دقت 99.69٪ رسید. در نهایت، مدل پیشنهادی **CNN–multi-head attention–LSTM** بهترین توازن میان دقت و پیچیدگی را فراهم کرده و با تنها 2.70 میلیون پارامتر به دقت 99.00٪ دست یافته است. تعداد پارامترهای هر مدل (قابل آموزش و غیرقابل آموزش) معیاری برای شفافیت در سنجش پیچیدگی محاسباتی و کارایی آن‌ها محسوب می‌شود. اطلاعات در جدول5 مقاله، قابل مشاهده است.

الگوریتم پیشنهادی برای **تشخیص پارکینسون** از سیگنال‌های صوتی خام، ترکیبی از **CNN**، **multi-head attention** و **LSTM** است.

1. **ورودی‌ها:**  داده‌های خام صوتی (هر نمونه در ابعاد 60×94×1).
2. **استخراج ویژگی‌ها:**  از هر سیگنال صوتی، **mel spectrogram** و **MFCC** استخراج می‌شوند تا هم اطلاعات فرکانسی و هم ویژگی‌های خاص گفتاری به دست بیاید.
3. **ترکیب ویژگی‌ها: دو ویژگی** mel spectrogram و MFCC کنار هم قرار می‌گیرند و یک نمایش ترکیبی (feature map) ساخته می‌شود.
4. **استخراج الگوهای مکانی :(Spatial features)** این نمایش ترکیبی از چندین لایه **Conv2D** عبور می‌کند.
   * در هر لایه **ReLU activation** و **L2 regularization** به‌کار می‌رود.
   * در لایه‌های زوج، علاوه بر این، **MaxPooling** برای کاهش ابعاد و **Dropout** برای جلوگیری از overfitting انجام می‌شود.
5. **:Flattening**خروجی لایه‌های کانولوشنی flat می‌شود تا آماده ورود به بخش بعدی شود.
6. **:Multi-Head Attention**روی ویژگی‌ها اعمال می‌شود تا مدل بتواند بخش‌های مهم‌تر دنباله صوتی را شناسایی کند و روابط پیچیده بین زمان–فرکانس را یاد بگیرد.
7. **:LSTM**داده‌ی پردازش‌شده توسط attention به لایه‌های **LSTM** داده می‌شود تا وابستگی‌های زمانی بلندمدت در گفتار ضبط شود. از Dropout هم در این مرحله برای جلوگیری از overfitting استفاده می‌شود.
8. **خروجی:** ویژگی‌های نهایی از طریق یک **Sigmoid activation** عبور می‌کنند تا مدل برچسب هر ورودی را پیش‌بینی کند.
9. **آموزش مدل**:
   * خطای پیش‌بینی با **binary cross-entropy loss** محاسبه می‌شود.
   * وزن‌های مدل با استفاده از **Adam optimizer** به‌روزرسانی می‌شوند.

تصویر زیر از مقاله، جزییات الگوریتم را به شکل شبه کد نشان می دهد.



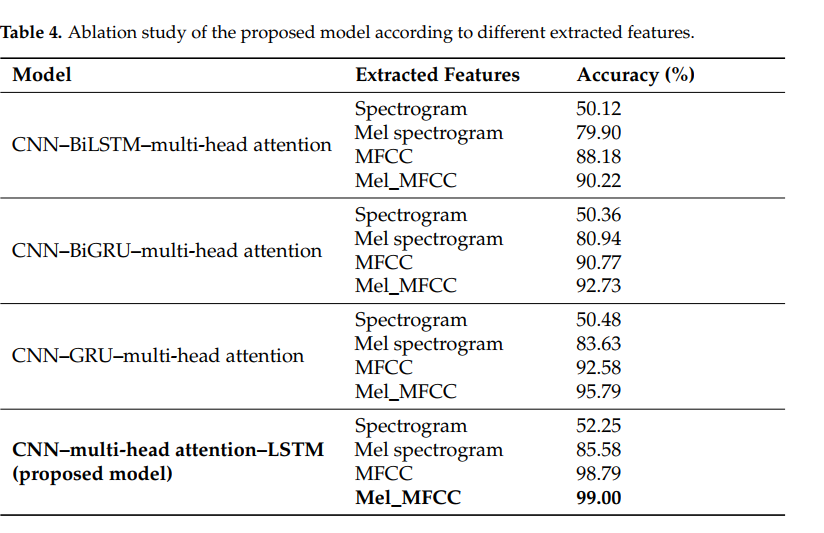
در نهایت، جدول بعدی یعنی جدول3 از مقاله، هایپرپارامترهای انتخابی مدل پیشنهادی را نشان می‌دهد.A white background with black text

AI-generated content may be incorrect.

## ویژگی ها

مدل پیشنهادی ورودی خودش را از روی **سیگنال‌های گفتار** (voice recordings) به‌دست می‌آورد. این ورودی‌ها بعد از پردازش اولیه به چهار نوع ویژگی اصلی تبدیل می‌شوند:

1. **Spectrogram**
   * صدای خام به قطعه‌های کوتاه تقسیم می‌شود و روی هر قطعه تحلیل فرکانسی انجام می‌گیرد.
   * خروجی یک ماتریس زمان–فرکانس است که در آن محور افقی نشان‌دهنده زمان و محور عمودی نشان‌دهنده فرکانس است.
   * این نمایش، تغییرات انرژی در باندهای فرکانسی مختلف را در طول زمان نشان می‌دهد.
   * برای بیماری پارکینسون، نوسانات ظریف در گفتار ممکن است در این نمایش قابل تشخیص باشند.
   * اما این ویژگی گاهی اطلاعات کلی و اضافی دارد و دقت تشخیص را محدود می‌کند (حدود 50–52%).
2. **Mel Spectrogram**
   * مشابه spectrogram است، با این تفاوت که فرکانس‌ها به مقیاس **Mel** نگاشت می‌شوند.
   * مقیاس Mel به گونه‌ای طراحی شده که بیشتر با نحوه شنیدن انسان سازگار باشد.
   * این ویژگی کمک می‌کند تغییرات فرکانسی که برای گوش انسان مهم‌ترند بهتر برجسته شوند.
   * برای داده‌های پارکینسون، mel spectrogram نسبت به spectrogram خام اطلاعات معنادارتری دارد (دقت حدود 79–85%).
3. **Mel-Frequency Cepstral Coefficients (MFCCs)**
   * ویژگی‌هایی هستند که از mel spectrogram استخراج می‌شوند.
   * این ویژگی‌ها الگوهای مربوط به تولید صوت توسط دستگاه گفتاری (مثل تارهای صوتی و حفره دهان) را خلاصه می‌کنند.
   * MFCCها بسیار حساس به تغییرات ریز در گفتار هستند، مانند **jitter** و تغییرات **pitch** که در بیماران پارکینسون دیده می‌شود.
   * به همین دلیل، استفاده از MFCC به دقت بالاتری می‌انجامد (حدود 88–98%).
4. **ترکیب Mel + MFCC (Mel\_MFCC)**
   * در این حالت، همزمان از mel spectrogram و MFCC استفاده می‌شود.
   * این ترکیب هم محتوای فرکانسی (mel spectrogram) و هم ویژگی‌های خاص گفتار (MFCC) را پوشش می‌دهد.
   * بهترین کارایی مدل در این حالت به دست آمده است و دقت تقریباً به 99% می‌رسد.

جدول زیر، کارآیی هر ورودی را در مدل پیشنهادی و مدل های پیشین نشان می دهد.

## داده

### دیتاست

دیتاست Italian Parkinson’s Voice and Speech یا به اختصار Italian در این گزارش، شامل 831 فایل صوتی شامل 336 نمونه از افراد سالم و 495 نمونه از بیماران پارکینسون است. داده‌ها در شرایط کنترل‌شده (محیط بدون بازتاب، دمای ثابت و فاصله مشخص از میکروفون) جمع‌آوری گردیدند. محتوای گفتار شامل متون متوازن از نظر واجی، هجای “pa–ta”، واکه‌های ساده a، e، i، o، u و مجموعه‌ای از کلمات و عبارات است. هرچند یک **عدم‌تعادل جزئی** بین داده‌های دو گروه وجود دارد (تعداد بیماران بیشتر است)، اما طراحی به گونه‌ای انجام شده تا تحلیل منصفانه و معتبر باقی بماند.

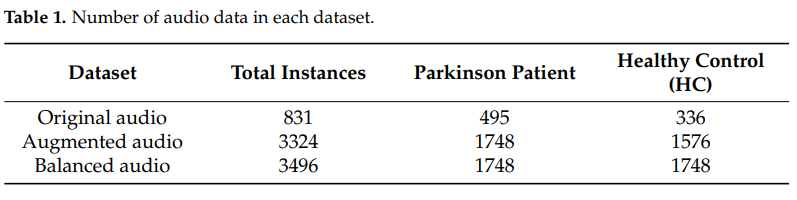
### آماده سازی داده

در این مرحله، ابتدا سیگنال‌های صوتی پیوسته به فریم‌های 30 میلی‌ثانیه‌ای تقسیم می‌شوند تا اطلاعات فرکانسی کوتاه‌مدت استخراج شود. سپس فایل‌های صوتی با نرخ نمونه‌برداری 16کیلوهرتز و طول 3 ثانیه، تحت تکنیک‌های مختلف **data augmentation** فرار می‌گیرند تا داده‌ها متنوع‌تر شوند و شرایط واقعی گفتار شبیه‌سازی گردد. این روش‌ها شامل:

* **:Pitch Scaling** تغییر ارتفاع صدا با ضریب 1.5 برای افزایش فرکانس بدون تغییر طول، که کمک می‌کند تغییرات ظریف در الگوهای صوتی مرتبط با پارکینسون بهتر شناسایی شوند
* **:Random Gain** تغییر تصادفی بلندی صدا بین 0.9 تا 1.1 برابر، برای شبیه‌سازی شرایط محیطی مختلف
* **:White Noise Addition**افزودن نویز سفید با ضریب 0.1 برای مقاوم‌سازی مدل در برابر نویز پس‌زمینه

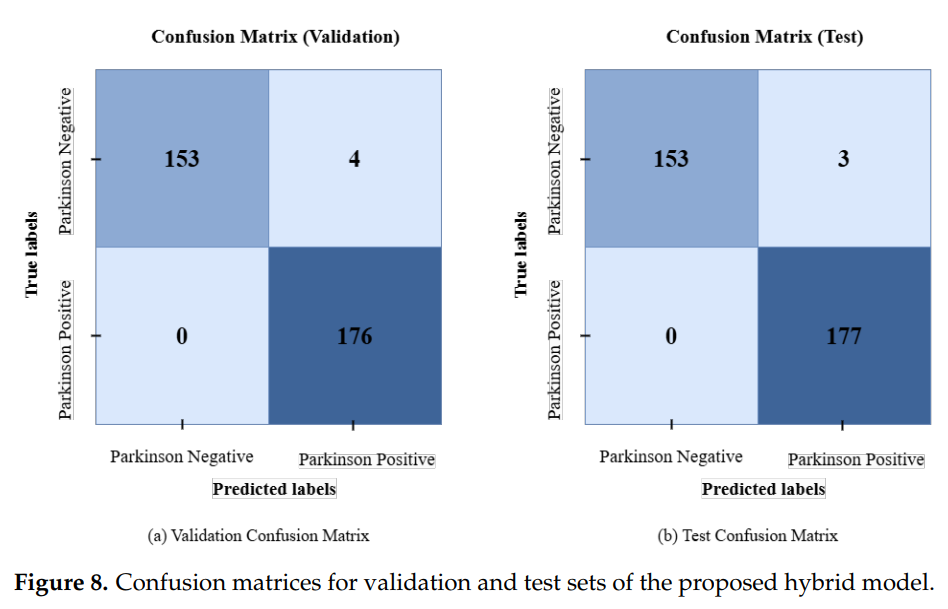
پس از این مرحله، برای رفع مشکل **عدم‌تعادل کلاسی** (تعداد کمتر نمونه‌های پارکینسون نسبت به سالم‌ها)، از **random oversampling** استفاده می‌شود. این روش با تکرار نمونه‌های کلاس اقلیت، داده‌ها را متعادل می‌کند و باعث می‌شود مدل توانایی یادگیری منصفانه و عملکرد بهتری در تشخیص هر دو گروه داشته باشد.

نتیجه ی این عملیات در جدول1 مقاله در زیر قابل مشاهده است.



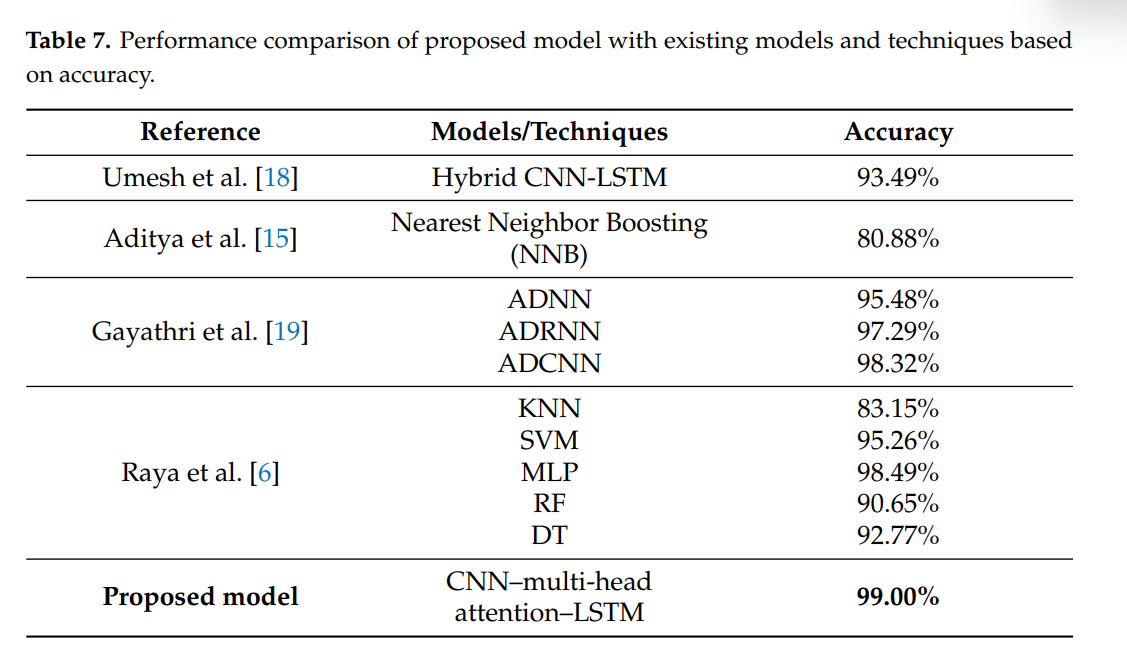
## نتایج

A graph with a line

AI-generated content may be incorrect.نتایج مدل پیشنهادی نشان می‌دهد که این روش در تشخیص افراد سالم از بیماران پارکینسون تقریباً بدون خطا عمل کرده است. همان‌طور که جدول 6 نشان می‌دهد، برای گروه سالم دقت 1.00 و یادآوری 0.98 به دست آمده و برای بیماران پارکینسون دقت 0.98 و یادآوری کامل 1.00 ثبت شده است؛ یعنی در هر دو کلاس، مدل به f1score برابر با 0.99 رسیده است. بررسی confusion matrixهم نشان می‌دهد که در مرحله اعتبارسنجی تنها چهار مورد خطای مثبت کاذب رخ داده و هیچ نمونه‌ای از بیماران از دست نرفته است؛ به عبارت دیگر، مدل توانسته 176 فرد سالم و 153 بیمار پارکینسون را درست تشخیص دهد. با این حال، در مرحله آزمون افت محسوسی مشاهده شد؛ به‌گونه‌ای که هرچند 153 فرد سالم درست شناسایی شدند، مدل در تشخیص بیماران پارکینسون ضعیف عمل کرده و سه منفی کاذب و 177 مثبت کاذب ایجاد کرده است. با وجود این، منحنی ROC با AUC برابر 1.00 نشان می‌دهد که این مدل پتانسیل بالایی برای جداسازی دو گروه دارد و در شرایط بهینه می‌تواند حساسیت بالا و نرخ خطای بسیار پایینی را تضمین کند.

### مقایسه با مدل های دیگر

مقایسه با مدل‌های موجود نشان می‌دهد که مدل ترکیبی پیشنهادی CNN–multi-head attention–LSTM با دقت 99.00٪ عملکرد بهتری نسبت به سایر روش‌ها دارد. این مدل از نظر دقت، روش‌هایی مانند Hybrid CNN-LSTM (93.49٪)، NNB (80.88٪)، و حتی مدل‌های قوی‌تر مانند ADCNN (98.32٪) و ADRNN (97.29٪) را پشت سر گذاشته است. تنها مدل MLP با دقت 98.49٪ عملکردی نزدیک به آن داشته است. دلیل برتری این مدل، ترکیب توانایی CNN در استخراج ویژگی‌های فضایی و LSTM در یادگیری وابستگی‌های زمانی، همراه با multi-head attention برای تمرکز بر بخش‌های کلیدی سیگنال است. علاوه بر این، به‌کارگیری تکنیک‌های regularization برای جلوگیری از overfitting و random oversampling برای رفع عدم‌توازن داده‌ها، موجب افزایش پایداری و قدرت پیش‌بینی مدل شده است. به‌طور کلی، نتایج نشان می‌دهد که رویکرد پیشنهادی توانسته است سطح جدیدی از دقت و قابلیت اطمینان در تشخیص پارکینسون ارائه دهد.



### بحث

این مطالعه نشان داد که معماری ترکیبی **CNN–multi-head attention–LSTM** می‌تواند با دقت 99٪ به‌عنوان روشی غیرتهاجمی برای شناسایی زودهنگام پارکینسون عمل کند. در آینده، پژوهش بر روی **مدالیته‌های دیگر** مانند دست‌خط و راه‌رفتن، استفاده از روش‌های **توضیح‌پذیری** (Grad-CAM, SHAP)، و توسعه‌ی **مدل‌های سبک و قابل‌استفاده در زمان واقعی** متمرکز خواهد بود. همچنین، گسترش دیتاست‌های چندحالته و توجه به **کاربرد عملی و دسترس‌پذیری**، گام‌های بعدی برای پیاده‌سازی واقعی این رویکرد هستند.

# مقاله ی Rahmatallah et al., 2025

در مطالعه‌ی Rahmatallah **و همکاران (2025)** تحت عنوان “*Pre-trained convolutional neural networks identify Parkinson’s disease from spectrogram images of voice samples*”، نویسندگان با بهره‌گیری از داده‌های صوتی بیماران پارکینسون و افراد سالم، رویکردی نوین برای تشخیص این بیماری ارائه کرده‌اند. این مقاله که در ژورنال **Scientific Reports** منتشر شده است.

این پژوهش به بررسی استفاده از شبکه‌های عصبی کانولوشنی (CNN) مبتنی بر انتقال یادگیری برای شناسایی بیماری پارکینسون از روی اسپکتروگرام صدای واکه‌ی کشیده /a/ پرداخته است. اهمیت این موضوع از آنجاست که حدود ۹۰٪ بیماران پارکینسون دچار مشکلات گفتاری هستند و تحلیل صوت می‌تواند روشی غیرتهاجمی، سریع و در دسترس برای تشخیص و پایش بیماری باشد.

برای انجام آزمایش‌ها، دو پایگاه داده به‌کار گرفته شد: یکی UAMS dataset شامل صدای بیماران و افراد سالم ضبط‌شده از طریق خطوط تلفنی آنالوگ با پهنای باند محدود، و دیگری mPower dataset شامل صدای ضبط‌شده با گوشی‌های هوشمند با کیفیت بالاتر. از هر دو مجموعه، اسپکتروگرام خطی و اسپکتروگرام مل-اسکیل (Mel-scale) تولید شد. مدل Inception V3 که پیش‌تر روی ImageNet آموزش دیده بود، با تغییر لایه‌های انتهایی برای دسته‌بندی بیماران پارکینسون (PwPD) و افراد سالم (HC) استفاده شد.

برای مقایسه عملکرد، روش‌های کلاسیک نیز به کار رفتند؛ شامل استخراج ویژگی‌های آکوستیکی (مانند jitter، shimmer و فرکانس پایه) و ویژگی‌های طیفی (مانند MFCC، LPCC) و سپس اعمال طبقه‌بندهای Logistic Regression و Random Forest که نتایج نشان داد که CNN مبتنی بر اسپکتروگرام‌ها به‌طور قابل‌توجهی عملکرد بهتری نسبت به این روش‌های سنتی دارد.

در بخش نتایج، شبکه CNN توانست با استفاده از اسپکتروگرام خطی به AUC حدود ۰٫۹۵ در مجموعه UAMS و ۰٫۹۲ در مجموعه mPower دست یابد. با استفاده از اسپکتروگرام‌های مل-اسکیل، این مقادیر به ترتیب به ۰٫۹۷ و ۰٫۹۵ افزایش یافتند که اگرچه بهبود اندکی است، اما از نظر آماری معنادار گزارش شد. نکته مهم اینکه حتی ضبط صدا با تلفن و کیفیت پایین نیز به کمک CNN توانست نتایج قابل اعتمادی ارائه دهد.

تحلیل ویژگی‌ها نشان داد که عوامل مهم در دو مجموعه داده متفاوت بودند؛ در مجموعه mPower، میانگین و انحراف معیار فرکانس پایه بیشترین اهمیت را داشتند، در حالی‌که در مجموعه UAMS ضرایب بالاتر MFCC و LPCC نقشی کلیدی داشتند. این تفاوت نشان می‌دهد که محیط ضبط صدا و نوع دستگاه می‌تواند بر ویژگی‌های قابل‌استخراج و در نتیجه دقت مدل‌ها اثر بگذارد.

## مدل

در این پژوهش دو رویکرد اصلی روش‌های سنتی مبتنی بر ویژگی و یادگیری عمیق مبتنی بر شبکه‌های عصبی کانولوشنی برای شناسایی بیماری پارکینسون از روی صدای واکه‌ی کشیده /a/ به‌کار گرفته شد.

برای طبقه‌بندی این بردارهای ویژگی از دو الگوریتم مرسوم استفاده شد Logistic Regression (LR) و Random Forest (RF) که مدل LR با انتخاب تدریجی متغیرها و اعتبارسنجی سه‌تایی بهینه‌سازی شد. در مقابل، Random Forest با الگوریتم Breiman پیاده‌سازی گردید، شامل ۱۰۰۰ درخت تصمیم، انتخاب تصادفی شش متغیر در هر گره، و حداقل اندازه پنج برای گره‌های نهایی. این تعداد بالای درخت‌ها موجب پایداری بیشتر شد، بدون آنکه پیچیدگی محاسباتی به شکل چشمگیری افزایش یابد.

در رویکرد دوم، از یک مدل یادگیری عمیق استفاده شد. بدین منظور شبکه Inception V3 که از قبل بر روی مجموعه داده بزرگ ImageNet آموزش دیده بود، به‌کار گرفته شد. بخش استخراج ویژگی این شبکه ثابت نگه داشته شد و لایه‌های نهایی با معماری جدید جایگزین شدند. معماری جدید شامل یک لایه Batch Normalization برای کاهش تغییرات آماری داده‌ها و تسریع یادگیری، دو لایه کاملاً متصل (Dense) با ۱۰۲۴ نورون و فعال‌ساز ReLU، و در نهایت یک لایه خروجی با دو نورون و فعال‌ساز Softmax برای دسته‌بندی بیماران و افراد سالم بود. برای آموزش، تنظیمات زیر اعمال شد: نرخ یادگیری اولیه ۰.۰۰۱ با بهینه‌ساز Adam، اندازه دسته (Batch size) برابر ۴، تعداد دوره‌های آموزشی (Epoch) محدود به ۱۰، و میزان Dropout برابر ۲۰٪ برای جلوگیری از بیش‌برازش. داده‌ها در قالب تصاویر اسپکتروگرام خطی و مل‌اسکیل ۱.۵ ثانیه‌ای با اندازه ۶۰۰×۶۰۰ پیکسل به مدل داده شدند. اعتبارسنجی نیز به روش Repeated Holdout انجام شد؛ یعنی داده‌ها در ۱۰۰ تکرار به‌طور تصادفی به نسبت ۷۰٪ برای آموزش و ۳۰٪ برای آزمون تقسیم شدند تا پایداری عملکرد بررسی گردد.

## ویژگی ها

در بخش روش‌های سنتی، ابتدا ویژگی‌های مختلف از سیگنال صوتی استخراج شدند. دسته اول ویژگی‌های آکوستیکی بودند که مستقیماً رفتار تارهای صوتی و کیفیت آوا را نشان می‌دهند. این دسته شامل فرکانس پایه، میانگین و انحراف معیار آن، فرمنت‌ها که بازتاب رزونانس مجرای صوتی هستند، نسبت هارمونیک به نویز (HNR) که بیانگر خشونت یا وضوح صداست، و معیارهای ناپایداری همچون Jitter (تغییرات نامنظم فرکانس پایه) و Shimmer (تغییرات نامنظم شدت صدا) بود. در مجموع ۲۳ شاخص آکوستیکی محاسبه شد.

دسته دوم ویژگی‌های طیفی بودند که به کمک تحلیل پنجره‌ای سیگنال استخراج شدند: ضرایب LPC، LAR، LPCC و MFCC که این ضرایب نمایانگر ساختار طیفی و جداسازی مؤلفه‌های منبع گلوتال و مجرای صوتی هستند. از هر مجموعه ضرایب، میانگین و واریانس محاسبه شد تا نماینده کل ضبط صوت باشد. در برخی موارد، ترکیب ویژگی‌های آکوستیکی و طیفی نیز به‌کار گرفته شد تا اطلاعات منبع و فیلتر همزمان لحاظ شود.

برای آماده‌سازی ورودی مدل، ابتدا فایل‌های صوتی خام به فرمت یکسان (WAV) تبدیل شدند و صداها در یک بازه‌ی مشخص نرمال‌سازی شدند. سکوت‌های ابتدای فایل و انتهای آن حذف گردید تا فقط بخش‌های حاوی صدا باقی بمانند. اگر بعد از این مرحله صدای باقی‌مانده کمتر از ۱.۵ ثانیه بود، آن نمونه کنار گذاشته شد. در نهایت از هر ضبط یک بازه‌ی ۱.۵ ثانیه‌ای انتخاب شد تا همه‌ی نمونه‌ها طول یکسان داشته باشند.

سپس دو نوع تصویر از این بازه‌ها ساخته شد:

* اسپکتروگرام خطی (Linear): در این حالت فرکانس‌ها به‌صورت مستقیم و ساده روی محور عمودی نمایش داده می‌شوند و تغییرات انرژی صدا در طول زمان روی تصویر مشخص است.
* اسپکتروگرام مل (Mel-scale): در این روش، مقیاس فرکانسی بر اساس نحوه‌ی شنیدن انسان تنظیم می‌شود؛ یعنی فرکانس‌های پایین با جزئیات بیشتری و فرکانس‌های بالا فشرده‌تر نمایش داده می‌شوند. این کار شباهت بیشتری به ادراک شنوایی ما دارد.

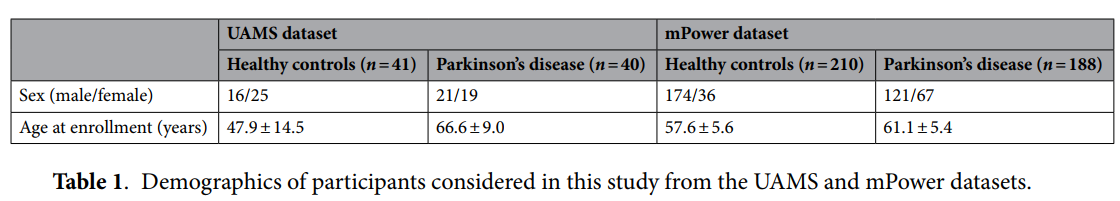
هر دو نوع اسپکتروگرام در قالب تصاویر رنگی با اندازه‌ی ۶۰۰ در ۶۰۰ پیکسل ذخیره شدند. بنابراین از هر نمونه‌ی صوتی دو تصویر ساخته شد: یکی خطی و دیگری مل. به این ترتیب، مدل می‌توانست هم از نمایش ساده فرکانس‌ها و هم از نمایش نزدیک‌تر به ادراک انسانی استفاده کند.

## دیتاست‌ها

برای ارزیابی مدل‌ها از دو مجموعه داده مستقل استفاده شد. نخست، UAMS dataset که در دانشگاه آرکانزاس گردآوری شده بود و شامل صدای بیماران پارکینسون و افراد سالم بود که از طریق تماس تلفنی و ضبط پیام صوتی در سامانه دیجیتال ذخیره شده بودند. این مجموعه به دلیل استفاده از خطوط تلفنی آنالوگ، تنها پهنای باند محدودی (۰.۳ تا ۳.۴ کیلوهرتز) را پوشش می‌داد و کیفیت صوت پایین‌تر بود.

دوم، mPower dataset که از طریق اپلیکیشن mPower روی گوشی‌های آیفون جمع‌آوری شد و ضبط صدا با کیفیت بالاتر و در محیط‌های متنوع انجام گرفته بود. برای اطمینان از دقت، فقط افراد ۵۰ تا ۷۰ ساله، بدون اختلالات دیگر مؤثر بر صدا، و با ضبط‌های فاقد نویز انتخاب شدند.

در نهایت، مجموعه UAMS شامل ۴۰ بیمار و ۴۱ فرد سالم و مجموعه mPower شامل ۱۸۸ بیمار و ۲۱۰ فرد سالم بود. جدول1 مقاله، جزییات این مجموعه ی انتخابی را نشان می دهد.



## نتیجه و مقایسه با مطالعات پیشین

نتایج نشان دادند که روش‌های سنتی مبتنی بر ویژگی، چه با Logistic Regression و چه با Random Forest، عملکرد محدودی داشتند (AUC حدود ۰.۶ تا ۰.۷۲) و تفاوت‌های زیادی میان دو مجموعه داده نشان دادند. در مقابل، مدل CNN مبتنی بر Inception V3 توانست به‌طور پایدار عملکرد بالایی ارائه دهد؛ در مجموعه UAMS با اسپکتروگرام خطی به AUC حدود ۰.۹۵ و با مل‌اسکیل به ۰.۹۷ رسید، و در مجموعه mPower نیز به ترتیب ۰.۹۲ و ۰.۹۵. این تفاوت هرچند کوچک بود، از نظر آماری معنادار محسوب شد و نشان داد که مل‌اسکیل مزیت نسبی دارد. نکته مهم آنکه CNN حتی در داده‌های تلفنی با کیفیت پایین نیز دقت بالایی داشت. مقایسه با مطالعات پیشین نشان می‌دهد که اگرچه کارهای قبلی هم بر برتری CNN در برابر روش‌های کلاسیک تأکید کرده بودند، اما بیشتر آن‌ها در شرایط آزمایشگاهی و با میکروفن‌های حرفه‌ای انجام شده بود. ارزش این پژوهش در آن است که اثبات کرد این رویکرد حتی در شرایط واقعی و ضبط‌های کم‌کیفیت نیز کارآمد است. از این‌رو، این روش می‌تواند پایه‌ای برای توسعه ابزارهای بالینی و سامانه‌های پایش از راه دور بیماران پارکینسون به‌ویژه در مناطق محروم باشد.

جزییات نتایج، در تصویر1 مقاله قابل مشاهده است.

