

Competição IEEE ICASSP

SAND (Speech Analysis for Neurodegenerative Diseases challenge)

Por que precisamos do SAND?

Este desafio decorre da necessidade de analisar biomarcadores não invasivos, objetivos e escaláveis, como sinais de fala, para diagnóstico precoce e monitoramento longitudinal de pacientes que sofrem de doenças neurodegenerativas. Isso ocorre porque doenças como a Esclerose Lateral Amiotrófica (ELA) apresentam desafios diagnósticos complexos devido a perfis de sintomas heterogêneos e características clínicas sobrepostas.

Limitações atuais

As ferramentas de diagnóstico atuais são amplamente baseadas em escalas clínicas subjetivas e frequentemente falham em detectar alterações precoces, resultando em intervenções tardias e cuidados abaixo do ideal para os pacientes. Isso ressalta a necessidade urgente do uso de biomarcadores não invasivos.

Faça parte da mudança!

Com este desafio, gostaríamos de redefinir a avaliação de doenças neurodegenerativas, posicionando a fala como um biomarcador central, alimentado por IA, para diagnóstico e monitoramento. Convidamos você a participar do desafio SAND com sua contribuição. As cinco equipes mais bem classificadas serão convidadas a apresentar seus trabalhos no IEEE ICASSP 2026. As contribuições aceitas serão publicadas nos anais oficiais do IEEE ICASSP (indexados pelo IEEE). A sessão dedicada ao desafio destacará as apresentações dos participantes com melhor desempenho e será encerrada com um painel de discussão. Não perca este evento envolvente!

Datas Importantes Todas as datas e horários especificados são baseados no fuso horário italiano, que é UTC+1 (Horário da Europa Central) durante o Horário Padrão e UTC+2 (Horário de Verão da Europa Central) durante o Horário de Verão.

- 1º de setembro de 2025: Abertura das inscrições para o desafio e liberação dos conjuntos de dados (treinamento)
- 1º de outubro de 2025: Liberação dos conjuntos de dados (teste) - primeiro dia para envio dos resultados e código
- 20 de novembro de 2025: Encerramento das inscrições para o desafio - último dia para inscrição, envio dos resultados e código
- 2 de dezembro de 2025: Anúncio dos resultados
- 7 de dezembro de 2025: Entrega dos artigos de 2 páginas (somente por convite)
- 11 de janeiro de 2026: Notificação de aceitação do artigo de 2 páginas
- 18 de janeiro de 2026: Entrega dos artigos de 2 páginas prontos para impressão

Nota: Cada fase começa às 00:00 (fuso horário italiano) do dia indicado acima.

Diretrizes para os participantes

- Os participantes podem enviar até três previsões por tarefa, sendo que apenas a submissão final será avaliada.

- As submissões podem ser feitas para uma ou ambas as tarefas. Cada submissão deve incluir uma breve descrição (2 páginas) da metodologia utilizada. Nesta fase, embora não seja um requisito obrigatório, é altamente recomendável que você redija o artigo de 2 páginas usando o modelo Word ou LaTeX fornecido no IEEE ICASSP PaperKit 2026.
- Cada participante pode fazer parte de apenas uma equipe.
- Cada equipe deve se registrar para participar deste desafio.
- Durante a inscrição, verifique atentamente a lista completa de membros da equipe, pois nenhuma modificação será permitida no futuro. Você não poderá adicionar membros adicionais ao artigo descritivo de 2 páginas, que será publicado nos anais do IEEE ICASSP, caso seja aceito.
- Cada equipe pode baixar o conjunto de dados de treinamento do painel.
- No momento apropriado (consulte as diretrizes), cada equipe pode baixar o conjunto de dados de teste do painel. Use-o para obter os resultados a serem enviados. Dados de teste não devem ser utilizados durante o treinamento. Todos os ajustes de parâmetros devem ser realizados utilizando o conjunto de treinamento, a partir do qual conjuntos de validação podem ser criados.
- Cada equipe pode enviar seus resultados a partir do painel.
- Envios incompletos ou com entradas ausentes serão excluídos da avaliação.
- A classificação final será publicada nesta plataforma e os vencedores serão notificados.
- A classificação será baseada no melhor desempenho alcançado para cada tarefa individual.
- Para determinar o(s) método(s) vencedor(es), os 3 modelos enviados com melhor desempenho (para cada tarefa) devem fornecer um arquivo executável (ou um notebook) para reproduzir os resultados da classificação/previsão. Recomendamos fortemente que todas as equipes divulguem seu código publicamente em seus perfis.
- As cinco melhores equipes serão selecionadas, levando em consideração a distribuição dos participantes entre as tarefas: as 2 melhores equipes da Tarefa 1 e as 2 melhores da Tarefa 2 serão escolhidas, com a quinta vaga indo para a terceira equipe classificada na tarefa com o maior número de envios.
- As cinco equipes com melhor classificação serão convidadas a enviar um artigo de duas páginas para ser apresentado no IEEE ICASSP 2026. O artigo de duas páginas deve ser formatado usando o modelo Word ou LaTeX fornecido no IEEE ICASSP PaperKit 2026 e enviado até o prazo final para publicação.
- Nossa implementação de linha de base inclui conjuntos de validação predefinidos (20% do conjunto de treinamento), que podem ser usados durante o treinamento. Detalhes da composição de nossos conjuntos de treinamento e validação de linha de base podem ser encontrados na seção Conjunto de Dados.
- O uso de recursos externos (dados de treinamento e/ou modelos pré-treinados) é permitido se os dados ou modelos pré-treinados forem de acesso público e gratuito, todos os recursos estiverem disponíveis antes do início do desafio e qualquer conjunto de dados ou modelo utilizado for devidamente citado na descrição de duas páginas da metodologia.
- Modelos desenvolvidos com dados privados não são permitidos.

Critérios de Avaliação (INCOMPLETO)

O desempenho dos modelos submetidos será avaliado com base na **pontuação média F1**, calculada usando o conjunto de testes de ensaios realizados. Para cada tarefa, essa métrica é calculada da seguinte forma:

$$\text{Avg. F1-Score} = \frac{1}{|C|} \sum_{c \in C} \frac{TP_c}{TP_c + \frac{1}{2}(FP_c + FN_c)}$$

onde TP_c é o número de verdadeiros positivos, FP_c de falsos positivos, FN_c de falsos negativos, todos relatados para a classe c , enquanto $|C|$ é 5 para a tarefa 1 e 4 para a tarefa 2.

A classificação final será baseada na pontuação média de F1, com pontuações mais altas indicando melhor desempenho de classificação/previsão. Escolhemos essa métrica por ser útil ao utilizar conjuntos de dados não balanceados. Em caso de empate, a originalidade da abordagem proposta servirá como critério adicional de avaliação. A decisão final caberá aos organizadores. Os participantes podem enviar resultados para uma ou ambas as tarefas. As cinco melhores equipes serão selecionadas com base no desempenho e na distribuição geral das submissões entre as tarefas (veja abaixo para esclarecimentos).

Nota: Os conjuntos de testes realizados para a Tarefa 1 e a Tarefa 2 diferem ligeiramente. **Certifique-se de usar o arquivo xls correto associado à tarefa correspondente da qual você está participando.**

Dataset

O conjunto de dados já está disponível para download no painel. Efetue login ou registre-se para acessar o painel e baixar o conjunto de dados. O conjunto de dados contém sinais de voz adquiridos de indivíduos adultos na faixa etária [18-90], coletados de 1º de janeiro de 2022 a 15 de junho de 2025. O estudo foi aprovado pelo Comitê de Ética do Hospital Universitário Federico II de Nápoles, Itália (ID do Protocolo: 100/17/ES01 e 93/2023). Detalhes da metodologia e dos métodos utilizados para conduzir o estudo estão disponíveis em <https://doi.org/10.1038/s41597-024-03597-2>. Em vez disso, informações sobre o aplicativo móvel Vox4Health, utilizado para adquirir os sinais de voz, estão disponíveis em https://doi.org/10.1007/978-3-319-40114-0_15. A equipe médica do Centro de ELA do Hospital Universitário Federico II de Nápoles realizou a avaliação clínica dos indivíduos e atribuiu a cada um deles uma pontuação ALSFRS-R, um valor inteiro de 0 a 4 para pacientes com ELA ou um valor de 5 para indivíduos saudáveis. Especificamos que nenhum indivíduo com ALSFRS-R igual a 0 (zero) está presente no conjunto de dados. A coleta consiste em 2.712 sinais de voz (relacionados a diferentes tarefas de fala) gravados de 339 falantes de italiano:

- 205 pacientes com ELA (121 homens e 84 mulheres) com diferentes graus de disartria.
- 134 indivíduos saudáveis (72 homens e 62 mulheres).

A distribuição por gênero dos pacientes nas duas tarefas é a seguinte. Para a tarefa 1, há 119 mulheres e 154 homens no conjunto de treinamento, enquanto há 27 mulheres e 40 homens no conjunto de teste. Para a tarefa 2, há 51 mulheres e 81 homens no conjunto de treinamento, enquanto há 14 mulheres e 19 homens no conjunto de teste. Todos os sinais foram registrados com o Vox4Health, um aplicativo de saúde móvel em um smartphone mantido a cerca de 20 centímetros da boca dos pacientes; o ângulo entre o celular e a boca era de cerca de 45 graus. Para o desafio SAND, o conjunto de dados foi dividido, mantendo um equilíbrio entre idade, gênero e gravidade da disartria (valor da escala ALSFRS-R). O conjunto de dados completo foi particionado em:

- 80% do conjunto de treinamento.
- 20% do conjunto de teste.

Para nossos experimentos com modelos de linha de base, usamos o conjunto de treinamento, que por sua vez foi dividido em:

- 80% para treinamento de linha de base.
- 20% para validação de linha de base.

As pontuações de linha de base foram calculadas em relação aos subconjuntos de validação. Os sujeitos utilizados nos conjuntos de treinamento e validação durante nossos experimentos para as linhas de base são relatados nos arquivos de dados, ou seja, `sand_task_1.xlsx` e `sand_task_2.xlsx`.

Arquivos de áudio

Para cada sujeito, este banco de dados contém:

- Gravação das vocalizações de cada vogal /a/, /e/, /i/, /o/ e /u/ por um mínimo de 5 segundos cada, garantindo uma intensidade sonora contínua; isso dá origem a cinco arquivos de áudio WAV.
- Gravações da voz do sujeito durante a repetição em uma única respiração de cada uma das três sílabas /pa/, /ta/ e /ka/ da maneira mais rápida possível; isso resulta em três arquivos de áudio WAV. Todas as gravações foram coletadas a uma frequência de amostragem de 8000 Hz e resolução de 16 bits. Elas ocorreram em condições de silêncio (< 30 dB de ruído de fundo), segura (taxa de umidade de cerca de 35-40%) e ausência de estresse emocional e fisiológico.

Arquivos de dados de treinamento

Os arquivos de dados de treinamento são `sand_task_1.xlsx` e `sand_task_2.xlsx`, respectivamente, para cada amostra da tarefa 1 e tarefa 2. Em detalhes, esses dois arquivos xlsx contêm as seguintes planilhas:

- **SAND - Conjunto de TREINAMENTO** - A Tarefa X contém a lista de indivíduos incluídos no conjunto de treinamento de desafio.
- **Linha de Base de Treinamento** - A Tarefa X contém a lista de indivíduos incluídos no conjunto de treinamento usado por nós para realizar a linha de base.
- **Linha de Base de Validação** - A Tarefa X contém a lista de indivíduos incluídos no conjunto de validação usado por nós para avaliar nossa linha de base. Onde X é 1 ou 2, dependendo da tarefa de interesse.

As planilhas dos arquivos xlsx contêm várias colunas, tanto recursos que podem ser usados como valores de entrada quanto rótulos de destino para serem usados como Verdade Fundamental durante a fase de treinamento. As colunas contidas em cada planilha variam dependendo da tarefa. Para a tarefa 1, cada planilha de `sand_task_1.xlsx` contém as seguintes colunas:

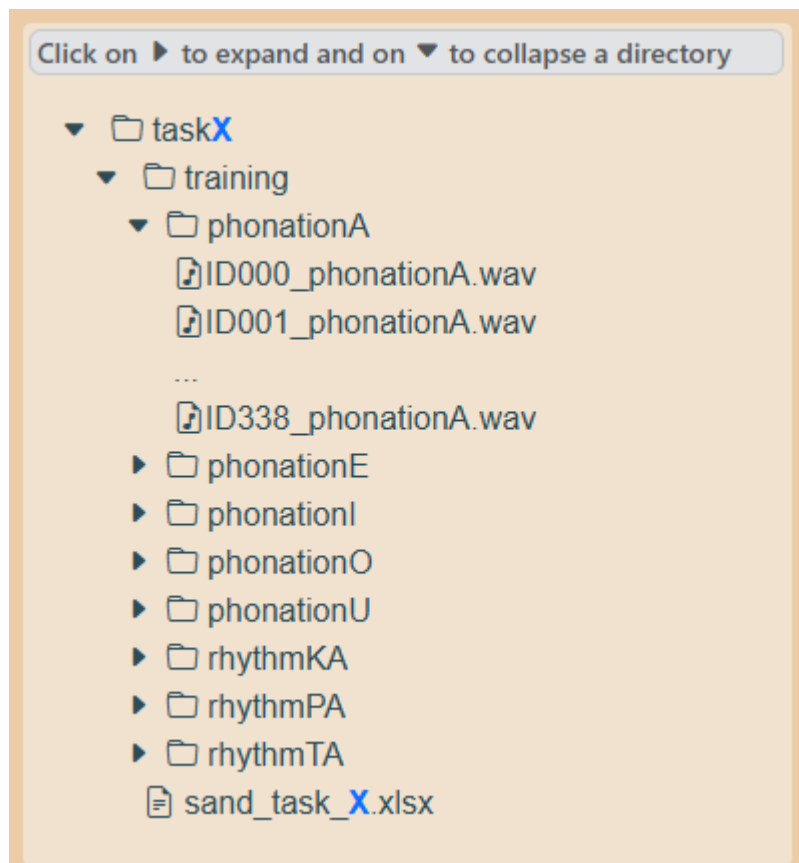
- ID: O identificador de cada sujeito, que corresponde ao prefixo de cada arquivo de áudio relacionado a esse sujeito específico.
- Idade: A idade do sujeito.
- Sexo: O gênero do sujeito.
- Classe: Rótulo Ground Truth, que representa a classe do sujeito. No caso de um sujeito com ELA, corresponde à gravidade do distúrbio de voz do paciente (disartria), ou seja, a pontuação da escala ALSFRS-R, que pode ter um valor de 1 a 4; enquanto, no caso de um sujeito saudável, é igual a 5.

Enquanto para a tarefa 2, cada planilha de `sand_task_2.xlsx` contém as seguintes colunas:

- ID: O identificador de cada sujeito, que corresponde ao prefixo de cada arquivo de áudio relacionado a esse sujeito específico.
- Idade: A idade do sujeito.

- Sexo: O gênero do sujeito.
- Meses: Número de meses entre a primeira e a última avaliação.
- ALSFRS--R_start: Gravidade do distúrbio vocal do paciente (disartria) na primeira avaliação, ou seja, a pontuação da escala ALSFRS-R, que pode ter um valor de 1 a 4.
- ALSFRS--R_end: Rótulo da verdade fundamental, a gravidade do distúrbio de voz do paciente (disartria) na última avaliação, ou seja, a pontuação da escala ALSFRS-R na última avaliação, que pode ter um valor de 1 a 4.

Nosso Conjunto de Dados de Treinamento é estruturado da seguinte forma:



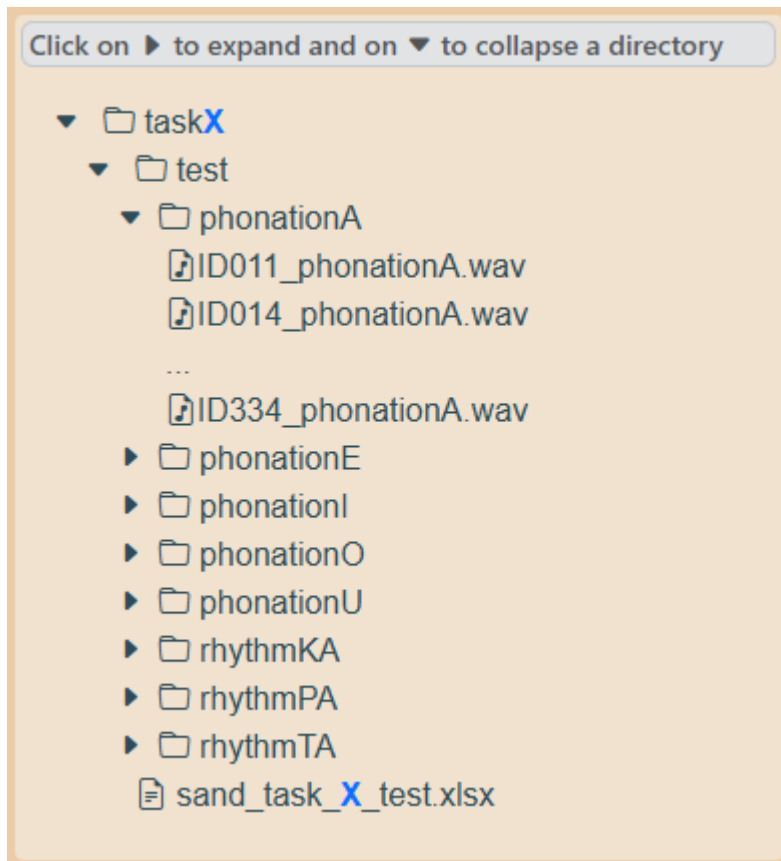
Neste caso também, **X** é 1 ou 2 dependendo da tarefa de interesse.

Arquivos de dados de teste

Os arquivos de dados de teste são `sand_task_1_test.xlsx` e `sand_task_2_test.xlsx`, respectivamente, para cada amostra de task1 e task2. A estrutura do arquivo é semelhante à do conjunto de dados de treinamento, exceto que, nesses arquivos de teste xlsx, a coluna de classe dos rótulos de verdade básica está vazia porque cada equipe deve estimar a classe usando seu próprio método.

Árvore de arquivos e diretórios (Conjunto de Dados de Teste)

Nosso Conjunto de Dados de Teste é estruturado da seguinte forma:



Neste caso também, X é 1 ou 2 dependendo da tarefa de interesse.

Tarefa 1 - Classificação Multiclasse

A Classificação Multiclasse no tempo 0 é proposta para identificar a abordagem mais confiável para detectar e classificar corretamente a gravidade dos distúrbios vocais (disartria), por meio da análise dos sinais de áudio, entre as cinco classes:

- ELA com disartria grave (Classe 1)
- ELA com disartria moderada (Classe 2)
- ELA com disartria leve (Classe 3)
- ELA sem disartria (Classe 4)
- Saudável (Classe 5)

A distribuição das classes para o conjunto de treinamento é a seguinte:

- Classe 1: 2,2%
- Classe 2: 9,55%
- Classe 3: 20,95%
- Classe 4: 27,94%
- Classe 5: 39,33%

Linha de base

Usamos um modelo Visual Transformer (ViT) como modelo de referência e obtivemos uma pontuação F1 média de 0,606 no conjunto de dados de validação obtido a partir do conjunto de dados de treinamento.

Tarefa 2 - Predição

A Tarefa 2 propõe prever a escala ALSFRS-R do paciente avaliado na última consulta de acompanhamento. Para esta tarefa, quatro classes são consideradas:

- ELA com disartria grave (Classe 1)
- ELA com disartria moderada (Classe 2)
- ELA com disartria leve (Classe 3)
- ELA sem disartria (Classe 4)

A distribuição das classes para o conjunto de treinamento é a seguinte:

- Classe 1: 13,64%
- Classe 2: 21,97%
- Classe 3: 28,79%
- Classe 4: 35,60%

O objetivo é prever com precisão a evolução da doença, ou seja, prever o valor da escala ALSFRS-R do paciente avaliado na última consulta de acompanhamento. Isso permitirá uma intervenção precoce e um melhor atendimento ao paciente.

Linha de base

Usamos um algoritmo de Árvore de Decisão Parcial (PART) como nosso modelo de linha de base e obtemos uma Pontuação F1 Média de 0,583 no conjunto de dados de validação obtido a partir do conjunto de dados de treinamento.