Revista de Casos e Consultoria, v. 17, n. 1, e42382, 2026
ISSN 2237-7417 | CC BY 4.0

Identificação automática da fala do magistrado em audiências trabalhistas utilizando aprendizado de máquina: contribuições para consultorias em inovação no Judiciário e para elaboração de casos de ensino

Automatic identification of judges' speech in labor court hearings using machine learning: contributions to innovation consulting in the Judiciary and to the development of teaching cases

Identificación automática del discurso de los jueces en las audiencias de los tribunales laborales mediante aprendizaje automático: contribuciones a la consultoría en innovación en el Poder Judicial y al desarrollo de casos didácticos

Recebido: 14/12/2025 | Revisado: 03/04/2026 | Aceito: 09/11/2026 | Publicado: 09/11/2026

Antonio Jhoseph Silva de Araújo
ORCID: https://orcid.org/0009-0006-0651-8400
Universidade Federal do Pará - UFPA, Brasil
E-mail: j.garibald@gmail.com

Adam Dreyton Ferreira dos Santos
ORCID: https://orcid.org/0000-0002-5940-4961
Universidade Federal do Sul e Sudeste do Pará - UNIFESSPA, Brasil
E-mail: adamdreyton@unifesspa.edu.br

Marcela Alves de Souza
ORCID: https://orcid.org/0000-0003-0241-1076
Universidade Federal do Sul e Sudeste do Pará - UNIFESSPA, Brasil
E-mail: marcela.alves@unifesspa.edu.br

Raphael Barros Teixeira
ORCID: https://orcid.org/0000-0003-2993-802X
Universidade Federal do Pará - UFPA, Brasil
E-mail: raphaelbt@ufpa.br

Resumo

A Justiça do Trabalho enfrenta desafios significativos relacionados à morosidade processual, especialmente na análise de gravações de audiências que subsidiam a elaboração de minutas e votos no

segundo grau. Nesse contexto, o uso de técnicas de Inteligência Artificial (IA) surge como alternativa para apoiar atividades de análise documental e reduzir o tempo empregado pelos servidores na identificação de informações relevantes. O Tribunal Regional do Trabalho da 8ª Região (TRT8) já utiliza a ferramenta 'Degrava' para transcrição automática de audiências, porém, à época desta pesquisa, ainda não dispunha de mecanismos capazes de identificar automaticamente o magistrado como orador. Assim, este estudo propõe e avalia um modelo de classificação de áudio para identificação da fala do magistrado em audiências trabalhistas. O *dataset* foi construído a partir de gravações públicas do TRT8, disponibilizadas no YouTube, especificamente da *playlist* "Sessões da 1ª Turma". Os áudios foram extraídos, segmentados, normalizados e processados por meio de uma metodologia automatizada desenvolvida nesta pesquisa, com extração de características acústicas baseada em *Mel-Frequency Cepstral Coefficients* (MFCCs). O conjunto de dados final contemplou cinco classes, cada uma correspondente a um magistrado. Foram avaliados diversos algoritmos de Aprendizado de Máquina, incluindo *Support Vector Machines* (SVM), *Logistic Regression* (LR), *Random Forest* (RF), AdaBoost e redes neurais do tipo *Multi Layer Perceptron* (MLP). Os resultados indicaram desempenho elevado para todos os modelos, com acurácia superior a 94%, destacando-se o MLP, que atingiu acurácia próxima de 99% e métricas equilibradas entre as classes. Os achados demonstram que modelos tradicionais, aliados a um pré-processamento consistente, são eficazes em cenários acústicos ruidosos, característicos das audiências trabalhistas. Conclui-se que a identificação automática da fala do magistrado é tecnicamente viável, compatível com a infraestrutura computacional do TRT8 e passível de integração ao ecossistema da ferramenta 'Degrava', promovendo ganhos de celeridade, segurança jurídica e potencial expansão para identificação de outros interlocutores. Além de sua contribuição técnica, o estudo apresenta utilidade prática para consultorias empresariais e organizacionais voltadas à inovação no setor público, especialmente em iniciativas de transformação digital, automação de fluxos e melhoria da eficiência institucional. Os achados também podem subsidiar a elaboração de casos de ensino sobre aplicação de inteligência artificial no contexto jurídico, permitindo seu uso por docentes e facilitadores em atividades de aprendizagem baseada em problemas e análise de casos.

Palavras-chave: Justiça do Trabalho; Aprendizado de Máquina; Identificação de Fala; MFCC; MLP.

Abstract

The Brazilian Labor Justice faces significant challenges related to procedural delays, particularly in the analysis of hearing recordings that support the preparation of draft decisions and opinions at the appellate level. In this context, the use of Artificial Intelligence (AI) techniques emerges as an alternative to support document analysis activities and reduce the time spent by court staff in identifying relevant information. The Regional Labor Court of the 8th Region (TRT8) already uses the Degrava tool for

automatic transcription of hearings; however, at the time of this research, it did not provide mechanisms capable of automatically identifying the judge as the speaker. Accordingly, this study proposes and evaluates an audio classification model for identifying judges' speech in labor court hearings. The dataset was built from public recordings made available by TRT8 on YouTube, specifically from the playlist "Sessões da 1ª Turma". The audio files were extracted, segmented, normalized, and processed using an automated methodology developed in this research, including acoustic feature extraction based on Mel-Frequency Cepstral Coefficients (MFCCs). The final dataset comprised five classes, each corresponding to a different judge. Several Machine Learning algorithms were evaluated, including Support Vector Machines (SVM), Logistic Regression, Random Forest, AdaBoost, and Multi-Layer Perceptron (MLP) neural networks. The results indicated high performance across all models, with accuracy above 94%, highlighting the MLP, which achieved accuracy close to 99% and balanced metrics among classes. The findings demonstrate that traditional models, combined with consistent preprocessing, are effective in noisy acoustic environments typical of labor court hearings. It is concluded that the automatic identification of judges' speech is technically feasible, compatible with the computational infrastructure available at TRT8, and suitable for integration into the Degrava ecosystem, providing gains in efficiency, legal certainty, and potential expansion to the identification of other interlocutors. Beyond its technical contribution, the study presents practical utility for business and organizational consultancies focused on innovation in the public sector, especially in initiatives for digital transformation, workflow automation, and improvement of institutional efficiency. The findings can also support the development of teaching cases on the application of artificial intelligence in the legal context, allowing its use by teachers and facilitators in problem-based learning activities and case analysis.

Keywords: Labor Justice; Machine Learning; Speech Identification; MFCC; MLP.

Resumen

La Justicia del Trabajo enfrenta desafíos significativos relacionados con la morosidad procesal, especialmente en el análisis de grabaciones de audiencias que sirven de base para la elaboración de borradores y votos en la segunda instancia. En este contexto, el uso de técnicas de Inteligencia Artificial (IA) surge como una alternativa para apoyar las actividades de análisis documental y reducir el tiempo empleado por los servidores en la identificación de información relevante. El Tribunal Regional del Trabajo de la 8ª Región (TRT8) ya utiliza la herramienta 'Degrava' para la transcripción automática de audiencias; sin embargo, en el momento de esta investigación, aún no contaba con mecanismos capaces de identificar automáticamente al magistrado como orador. En este sentido, este estudio propone y evalúa un modelo de clasificación de audio para la identificación del habla del magistrado en audiencias laborales. El dataset fue construido a partir de grabaciones públicas del TRT8 disponibles en YouTube,

específicamente de la lista de reproducción “Sessões da 1ª Turma”. Los audios fueron extraídos, segmentados, normalizados y procesados mediante una metodología automatizada desarrollada en esta investigación, incluyendo la extracción de características acústicas basada en Mel-Frequency Cepstral Coefficients (MFCCs). El conjunto de datos final contempló cinco clases, cada una correspondiente a un magistrado. Se evaluaron diversos algoritmos de Aprendizaje Automático, incluyendo Support Vector Machines (SVM), Regresión Logística, Random Forest, AdaBoost y redes neuronales de tipo Multi-Layer Perceptron (MLP). Los resultados indicaron un alto desempeño en todos los modelos, con una exactitud superior al 94%, destacándose el MLP, que alcanzó una exactitud cercana al 99% y métricas equilibradas entre las clases. Los hallazgos demuestran que los modelos tradicionales, combinados con un preprocesamiento consistente, son eficaces en entornos acústicos ruidosos, característicos de las audiencias laborales. Se concluye que la identificación automática del habla del magistrado es técnicamente viable, compatible con la infraestructura computacional del TRT8 y susceptible de integración en el ecosistema de la herramienta ‘Degrava’, promoviendo mejoras en la celeridad, la seguridad jurídica y el potencial de expansión para la identificación de otros interlocutores. Más allá de su contribución técnica, el estudio ofrece utilidad práctica para las consultorías empresariales y organizativas centradas en la innovación en el sector público, especialmente en iniciativas de transformación digital, automatización de flujos de trabajo y mejora de la eficiencia institucional. Los resultados también pueden servir de base para el desarrollo de casos de enseñanza sobre la aplicación de la inteligencia artificial en el ámbito jurídico, permitiendo su uso por parte de docentes y facilitadores en actividades de aprendizaje basado en problemas y análisis de casos.

Palabras clave: Justicia Laboral; Aprendizaje Automático; Identificación del Habla; MFCC; MLP.

Introdução

Um dos grandes problemas da justiça brasileira é a morosidade, despertando na sociedade a necessidade de se cobrar por uma justiça célere, capaz de entregar serviços efetivos a custos satisfatórios (STAATS; BOWLER; HISKEY, 2005). Segundo o relatório anual Justiça em Números 2024, ano-base 2023, disponibilizado pelo Conselho Nacional de Justiça (CNJ), foram registrados 35,3 milhões de novos processos. Ao final de 2023, havia 83,8 milhões de processos em curso; desse total, 18,5 milhões estavam suspensos.

Na Justiça do Trabalho em 2023, ramo com maior aumento percentual comparado ao ano de 2022, foram registrados aproximadamente 4,2 milhões de novos processos, o que representa um crescimento aproximado de 28,7% sobre 2022. Além disso, vale ressaltar que o tempo médio de tramitação dos processos pendentes está acima de 3 anos, e de processos baixados, resolvidos, está próximo a 2 anos

(CNJ, 2024).

De modo a mitigar essa problemática, alguns tribunais brasileiros já estão desenvolvendo soluções baseadas em Inteligência Artificial (IA). Dentre alguns projetos desenvolvidos e em produção, temos o VICTOR, um robô que auxilia na resolução de cerca de 10 mil recursos extraordinários que chegam ao Supremo Tribunal Federal (STF): uma tarefa que demandava em média 30 minutos de serviço é realizada em apenas cinco segundos (INAZAWA et al., 2019).

No Tribunal Regional do Trabalho da 8ª Região (TRT8), foi desenvolvida uma ferramenta de degravação de audiência, Degrava[1], com o objetivo de acelerar e facilitar a análise processual do Segundo Grau. Outros tribunais já implementaram soluções semelhantes, como o Tribunal Regional do Trabalho da 7ª Região (TRT7) - Ceará, que fez uso da digitação por voz presente no Google Docs, e o Tribunal de Justiça do Estado de Minas Gerais, que desenvolveu sua ferramenta de degravação chamada DORA.

Segundo dados do painel de monitoramento e-Gestão 2º Grau do TRT8 (TRT8, 2025), no ano de 2024 foram registrados aproximadamente 22,5 mil casos novos para análise, com aproximadamente 16,7 mil recursos ordinários. Até setembro de 2025, já foram registrados mais de 16 mil casos novos, destes, mais de 12 mil casos são de recursos ordinários. Estes processos, quando não têm sua ata disponível, precisam ter sua gravação revisada para subsidiar a elaboração da minuta de voto. Uma audiência tem uma média de duração de 30 minutos, isso resulta em aproximadamente 8,4 mil horas de audiência para serem analisadas no ano de 2024 e 6 mil horas no ano de 2025.

Com a implantação da ferramenta, não há mais a necessidade de os analistas assistirem a toda a gravação da audiência para realização de suas atividades, tornando a análise mais eficiente, já que é possível realizar buscas por palavras-chave e navegar para os pontos mais relevantes da audiência. Entretanto, tais ferramentas, à época desta pesquisa, ainda não ofereciam a funcionalidade de identificação automática do orador.

A presente pesquisa se justifica por sua relevância prática, científica e institucional. Do ponto de vista prático, o modelo sugerido ataca a morosidade processual na Justiça do Trabalho, um dos grandes desafios do judiciário brasileiro. Ao identificar automaticamente a fala do magistrado em audiências, a ferramenta permite que analistas e assessores localizem com celeridade os pontos decisórios da audiência, otimizando a elaboração de minutas de votos e aumentando a segurança jurídica ao mitigar erros de atribuição na transcrição.

Cientificamente, este trabalho ajuda a investigar a aplicação dos modelos de aprendizado de máquina em um cenário acusticamente adverso, o das audiências trabalhistas reais, oferecendo uma solução viável para instituições com recursos computacionais limitados.

Por fim, sua relevância institucional reside em complementar e potencializar a ferramenta de

[1] https://intranet.trt8.jus.br/noticia/trt-8-implementa-sistema-degrava-para-otimizar-analise-processual

degravação “Degrava” já em uso no TRT8.

Para além da contribuição científica e institucional, este estudo possui aplicabilidade prática em contextos de consultoria e ensino. No campo da consultoria, os resultados podem apoiar diagnósticos organizacionais e a formulação de recomendações voltadas à automação de processos, melhoria de fluxos de trabalho e modernização de serviços intensivos em análise de áudio e informação, à semelhança de estudos publicados na Revista de Casos e Consultoria que empregam métodos quantitativos e tecnológicos como ferramentas de consultoria organizacional (SILVA NETO; SOUZA, 2026) e como subsídio à transformação digital (OLIVEIRA et al., 2026). No campo educacional, o artigo pode servir como base para a construção de casos de ensino sobre transformação digital no Judiciário, uso de inteligência artificial em organizações públicas e tomada de decisão baseada em evidências, oferecendo subsídios para docentes e estudantes analisarem problemas reais e proporem soluções contextualizadas (SANTOS; AMORIM, 2026).

Metodologia

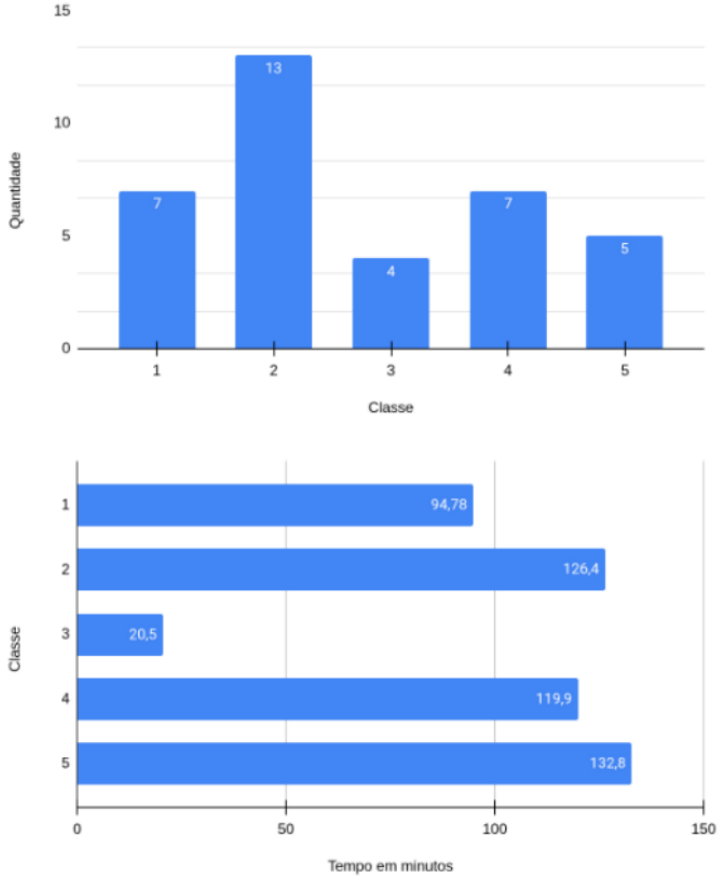
Dataset

Os dados utilizados para treinamento do modelo foram obtidos através do canal do TRT8 no YouTube, mais especificamente na *playlist* “Sessões da 1ª Turma”[2]. Todas as informações coletadas são de domínio público e de fácil acesso, conforme disposto na Lei nº 12.527/2011. Foram selecionados alguns vídeos para que pudessem ser realizadas as coletas dos áudios dos magistrados desta turma, de modo a gerar o *dataset* necessário para treinamento do modelo.

Neste trabalho, os desembargadores serão representados pelas classes de 1 a 5, respectivamente. Durante o processo de seleção dos vídeos, o critério da diversidade de ambiente no qual o magistrado estava inserido foi adotado visando abranger diferentes qualidades de áudio, proporcionando uma maior diversidade dos dados a fim de aprimorar a acurácia do modelo. A Figura 1 apresenta uma representação gráfica do quantitativo de vídeos selecionados e o tempo de fala de cada magistrado coletado.

[2] https://www.youtube.com/watch?v=Dvhi8CY5Zjo&list=PLuJHLZDF7t9ZQkZQ7si6a2_YGTEatIpcu

Figura 1 – Gráficos de quantidade de vídeos e tempo de falas coletadas



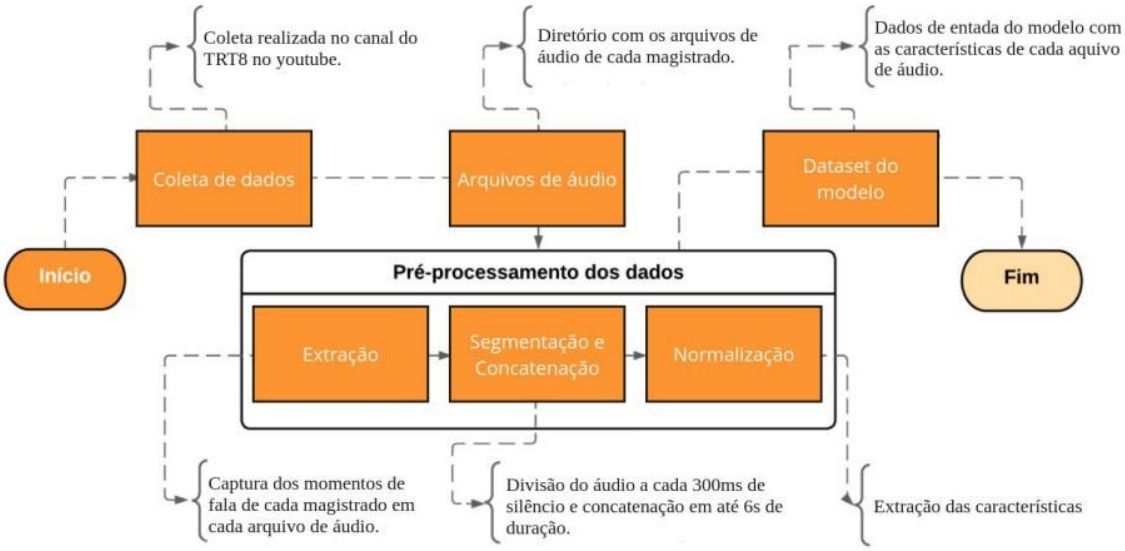
Fonte: Autores (2025).

Importante observar que a coleta de dados da "classe 3" ficou abaixo dos demais. Este fato justifica-se pela pouca quantidade de vídeos nos quais a desembargadora esteve presente, devido ao seu pouco tempo de exercício no cargo. No momento da coleta dos dados, ela havia recém-tomado posse.

A Figura 2 apresenta uma representação do processo de tratamento dos dados brutos para transformação nos dados de entrada dos algoritmos a serem testados. A fase subsequente à coleta de dados consistiu na extração e segmentação das intervenções individuais de cada magistrado. Considerando que a entrada do modelo proposto corresponde diretamente ao áudio extraído dos vídeos das audiências, tornou-se necessário automatizar o processo de extração e segmentação das falas, utilizando a linguagem Python[3] e a biblioteca *pydub*[4] (versão 0.25.1).

[3] https://www.python.org/ftp/python/3.12.10/python-3.12.10-amd64.exe
[4] https://pypi.org/project/pydub/

Figura 2 – Processo de tratamento dos dados

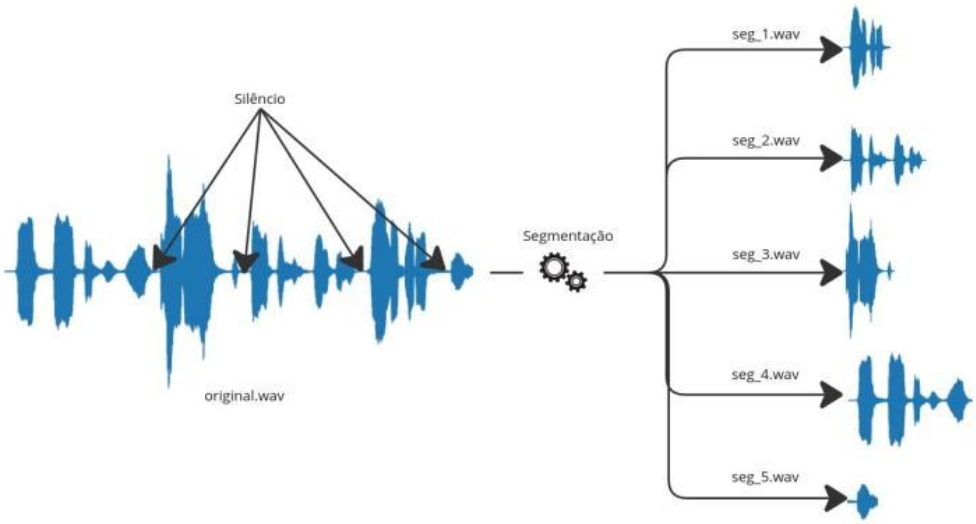


Fonte: Autores (2025).

Inicialmente, procedeu-se à identificação precisa dos trechos de fala de cada magistrado, o que resultou na elaboração de um arquivo no formato CSV. Este arquivo contém as informações para a delimitação temporal de cada intervenção, como o tempo de início ("*start*"), o tempo de término ("*end*") e o código de identificação do arquivo de áudio ("*code*"). A partir desse mapeamento, tornou-se possível a extração automatizada das falas.

A Figura 3 apresenta uma visão macro do processo de segmentação do áudio, realizada após a etapa de extração, que consiste na divisão do sinal a cada detecção de silêncio com duração igual ou superior a 300 milissegundos. Esse valor, definido empiricamente, visa otimizar o processo de análise ao remover trechos de silêncio irrelevantes, o que contribui para a redução da dimensionalidade dos dados, a diminuição da complexidade computacional e a otimização do tempo de treinamento do modelo.

Figura 3 – Segmentação do áudio

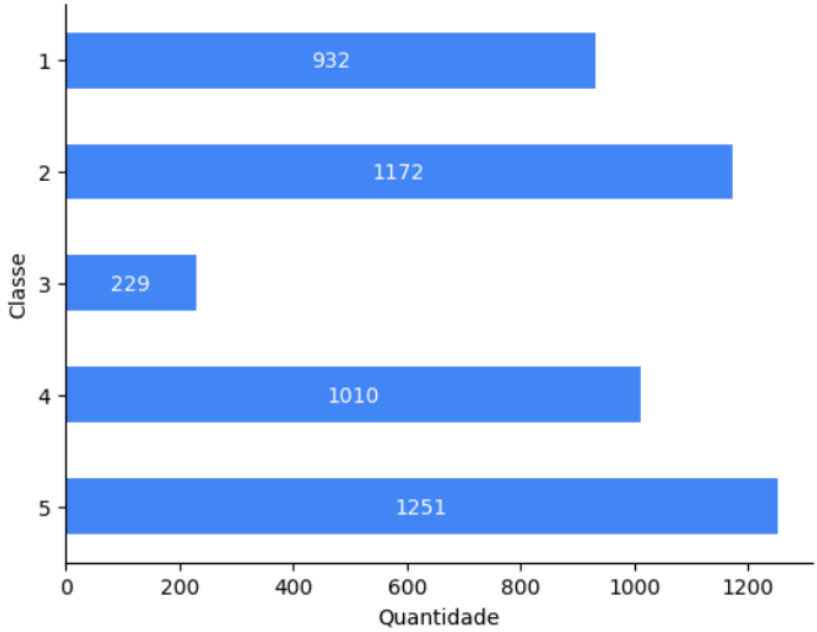


Fonte: Autores (2025).

A próxima etapa de pré-processamento consiste na concatenação dos arquivos de áudio, com o

objetivo de aprimorar a qualidade dos dados para o treinamento do modelo. A duração dos clipes concatenados, definida entre 4 e 6 segundos, baseia-se no estudo de Chu, Narayanan e Kuo (2009), que demonstrou a suficiência desse intervalo para a identificação de sons ambientais por ouvintes humanos com uma taxa de acerto de 82%. A Figura 4 apresenta a representação do quantitativo de arquivos gerados para cada classe.

Figura 4 – Quantidade de arquivos por classe após o processamento



Fonte: Autores (2025).

A etapa final de pré-processamento, denominada "Normalização" e apresentada na Figura 2, consiste em processar cada arquivo de áudio, extraindo as características MFCCs com número de coeficientes igual a 40. Para extração das características, foi utilizada a biblioteca *librosa* (versão 0.12.0), usando o interpolador *kaiser_fast*. Essas características serão utilizadas como atributos ("*features*") para o modelo de classificação.

A escolha de 40 coeficientes justifica-se pela necessidade de reter detalhes específicos do som que definem sua qualidade e origem. De acordo com Ganchev, Fakotakis e Kokkinakis (2005), o uso tradicional de apenas 13 coeficientes simplifica excessivamente o sinal, sendo útil para transcrição de fala, mas insuficiente para capturar a complexidade necessária em análises de identidade ou estilo.

Treinamento do Modelo de Classificação

Com a coleção de dados final definida, o próximo passo consiste em desenvolver e validar um modelo de classificação de áudio. Este modelo supervisionado exige uma amostra de dados rotulados para treinamento. Dessa forma, o objetivo do modelo é inferir qual magistrado está falando com base em um arquivo de áudio.

O treinamento foi conduzido utilizando algoritmos de aprendizado de máquina tradicionais, como *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Random Forest* (RF), *AdaBoost*, além do

Multi Layer Perceptron (MLP). Os dados provenientes da etapa anterior foram divididos na proporção 80/20, sendo 80% destinados ao treinamento e 20% destinados aos testes.

Visando garantir a reprodutibilidade dos experimentos e a transparência metodológica, foram estabelecidas configurações específicas de hiperparâmetros para cada classificador avaliado. A escolha desses parâmetros baseou-se em testes empíricos preliminares e recomendações da literatura para tarefas de classificação de áudio. A Figura 5 detalha as configurações finais adotadas para cada algoritmo no momento do treinamento. Os parâmetros não listados explicitamente foram mantidos com os valores padrão da biblioteca *scikit-learn*.

Figura 5 – Hiperparâmetros utilizados na configuração dos modelos

Algoritmo	Configuração dos Hiperparâmetros
SVM	*Kernel*: *RBF*; *C*: 1.0; *Gamma*: *scale*; *Probability*: *True*; *Shrinking*: *True*; *Cache Size*: 500 MB
MLP	*Hidden Layers*: (128, 64); *Activation*: *ReLU*; *Solver*: *Adam*; *Learning Rate*: 1e-3
LR	*Solver*: *lbfgs*; *Penalty*: *l2*; *Max Iter*: 1000; *C*: 1.0; *Fit Intercept*: *True*; *Multi-Class*: *auto*
RF	*N Estimators*: 100; *Criterion*: *gini*; *Max Depth*: *None*; *Min Samples Split*: 2; *Min Samples Leaf*: 1; *Max Features*: *sqrt*; *Bootstrap*: *True*
AdaBoost	*N Estimators*: 50; *Learning Rate*: 1.0; *Algorithm*: *SAMME.R*; *Base Estimator*: *Decision Tree* (*max_depth* = 1)

Fonte: Autores (2025).

A Figura 6 apresenta as médias das métricas obtidas para cada modelo considerando *precision*, *recall*, *f1-score* (*macro average*) e acurácia. De modo geral, os modelos apresentaram desempenho elevado, com médias superiores a 94% em todas as métricas avaliadas. O MLP e o SVM se destacaram como as abordagens mais eficazes, alcançando valores próximos ou superiores a 98% em *precision*, *recall* e *f1-score*, evidenciando alta capacidade de generalização e equilíbrio entre as classes. Modelos lineares, como LR, e métodos baseados em árvores, como RF, também apresentaram resultados consistentes, mantendo acurácias próximas a 97%. Por outro lado, o AdaBoost apresentou desempenho inferior, especialmente em termos de revocação, o que indica maior dificuldade na identificação completa de todas as instâncias positivas.

Figura 6 – Métricas dos modelos

Métrica	SVM	MLP	LR	RF	AdaBoost
Precision (macro avg)	0.989	**0.990**	0.980	**0.990**	0.910
Recall (macro avg)	0.986	**0.990**	0.982	0.934	0.833
F1-score (macro avg)	0.988	**0.990**	0.981	0.961	0.855
Acurácia	0.987	**0.988**	0.979	0.949	0.904

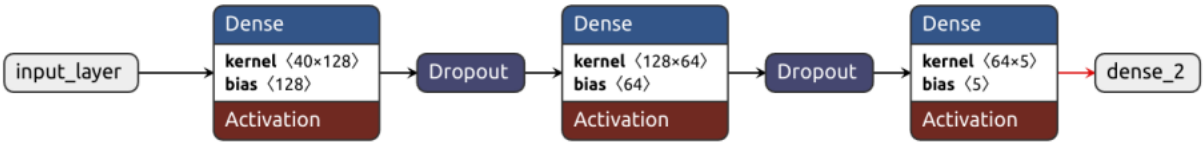
Fonte: Autores (2025).

Considerando todas as métricas, o MLP se apresenta como o melhor modelo, pois não apenas obteve a maior acurácia (0,988), mas também manteve consistência nas demais métricas e desempenho uniforme em todas as classes, inclusive naquelas com menor número de amostras. Esses resultados

reforçam a robustez das abordagens baseadas em redes neurais e máquinas de vetores de suporte para o problema em questão, mantendo um desempenho consistente em diferentes métricas de avaliação.

Conforme observado na Figura 7, a arquitetura do modelo MLP é composta por uma camada de entrada com 40 neurônios, correspondente aos 40 coeficientes MFCCs extraídos de cada sinal de áudio. Essa camada é seguida por três camadas densas totalmente conectadas, que aumentam gradualmente a capacidade do modelo de aprender representações complexas dos dados. A primeira camada densa possui 128 neurônios, seguida por uma camada *dropout* que desativa aleatoriamente 30% dos neurônios durante o treinamento. Isso é feito para forçar o modelo a não depender excessivamente de quaisquer combinações específicas de características, aumentando assim sua capacidade de generalizar para dados novos.

Figura 7 – Arquitetura do modelo MLP



Fonte: Autores (2025).

O treinamento do modelo foi realizado com 100 épocas utilizando o otimizador *Adam* (KINGMA; BA, 2014), que é amplamente reconhecido por sua eficiência em encontrar mínimos locais e acelerar a convergência, combinado com a função de perda *categorical cross-entropy*, que é apropriada para problemas de classificação multiclasse.

Além disso, aplicou-se a técnica de validação cruzada com 5 partições (*5-fold cross validation*), que divide o conjunto de dados em múltiplos subconjuntos para garantir que o modelo seja avaliado em diferentes amostras dos dados, reduzindo a variação entre as divisões e assegurando uma avaliação robusta e confiável do desempenho do modelo ao longo dos diferentes *folds* (BERRAR, 2019).

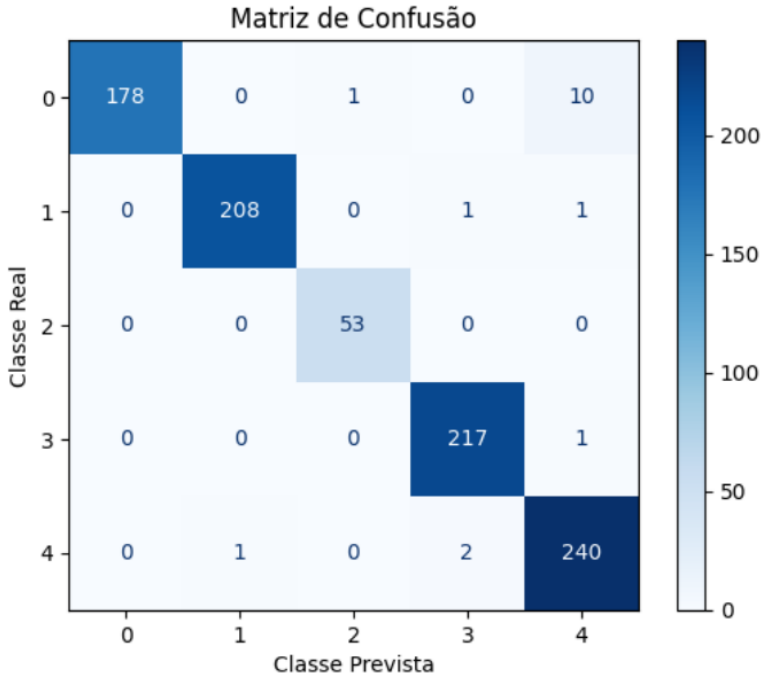
Quanto ao uso de ferramentas de Inteligência Artificial, durante o desenvolvimento desta pesquisa foram utilizadas ferramentas de Inteligência Artificial generativa, em especial o ChatGPT e o Gemini, como apoio complementar em atividades relacionadas à elaboração do manuscrito. Essas ferramentas auxiliaram em tarefas de revisão textual, correção gramatical e ortográfica, aprimoramento da clareza da redação, organização de ideias e sistematização preliminar de informações. Na etapa final de adequação ao periódico, utilizou-se também o Claude (Anthropic) como apoio à revisão ortográfica, à verificação das referências e à formatação do texto no modelo da revista. Ressalta-se, entretanto, que a utilização dessas ferramentas ocorreu apenas como suporte instrumental. A concepção da pesquisa, a definição da metodologia, a coleta e o tratamento dos dados, a análise dos resultados e a elaboração das conclusões permaneceram sob responsabilidade dos autores, que realizaram a validação crítica e científica de todo o conteúdo apresentado.

Resultados

Conforme apresentado na metodologia, o MLP se apresenta como o melhor modelo, pois não apenas obteve a maior acurácia (0,99), mas também manteve consistência nas demais métricas e desempenho uniforme em todas as classes, inclusive naquelas com menor número de amostras. A matriz de confusão apresentada na Figura 8 indica o desempenho deste modelo em identificar corretamente cada uma das classes de áudio. Observa-se uma alta taxa de acertos nas principais classes, com valores elevados na diagonal principal, o que reflete uma classificação precisa na maioria dos casos.

As confusões são mínimas, sugerindo que o modelo possui uma boa capacidade discriminatória. Algumas confusões pequenas ocorrem, como na Classe 0 com a Classe 4, em que 10 instâncias foram classificadas incorretamente. No entanto, esses erros são poucos em comparação com o total, indicando um bom desempenho geral.

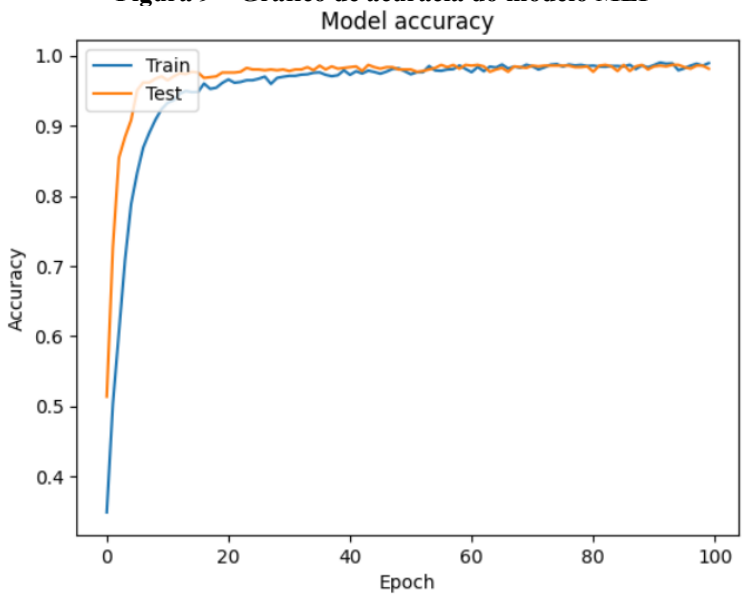
Figura 8 – Matriz de confusão do modelo MLP



Fonte: Autores (2025).

Na Figura 9, o gráfico de acurácia ao longo das épocas mostra uma rápida convergência do modelo, em que tanto o conjunto de treino quanto o de teste alcançam valores próximos de 1.0 (100%) de acurácia a partir de aproximadamente 20 épocas. A estabilidade entre as linhas de treino e teste indica que o modelo não apresenta sinais de *overfitting* ou *underfitting* significativos, mantendo uma performance consistente entre os dados vistos durante o treino e aqueles usados para teste. Esse comportamento é ideal, pois sugere que o modelo generaliza bem.

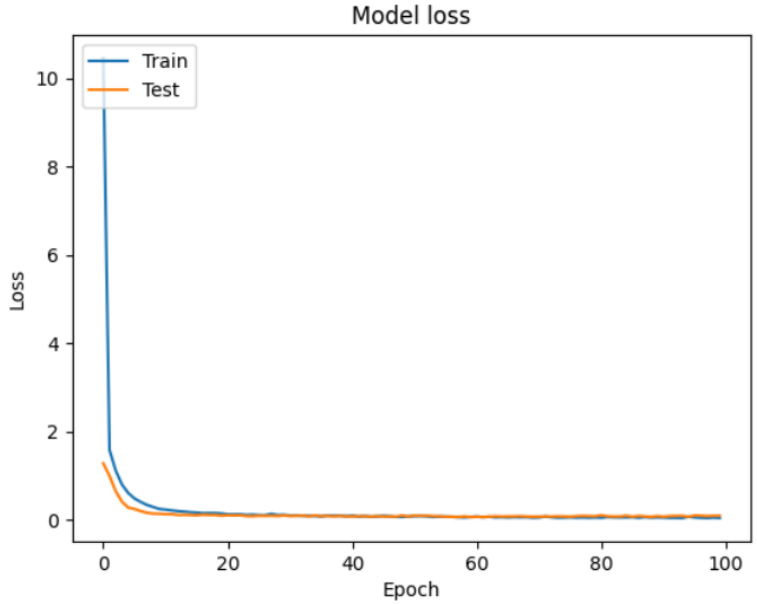
Figura 9 – Gráfico de acurácia do modelo MLP



Fonte: Autores (2025).

O gráfico de perda do modelo na Figura 10 reflete um comportamento de convergência semelhante ao da acurácia, com uma queda acentuada nas primeiras épocas e estabilização em valores próximos de zero para treino e teste após aproximadamente 20 épocas. Isso indica que o modelo minimizou bem a função de perda, alcançando um ajuste adequado aos dados. A baixa perda nos conjuntos de treino e teste, sem discrepâncias significativas, reforça a ideia de um modelo bem treinado, capaz de manter a generalização.

Figura 10 – Gráfico de perda do MLP

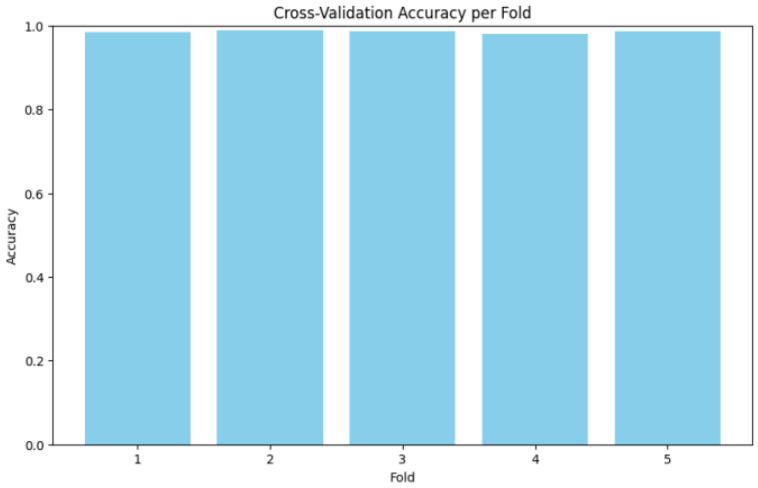


Fonte: Autores (2025).

O gráfico de acurácia por *fold* na validação cruzada, Figura 11, apresenta uma acurácia consistente em todos os cinco *folds*, com valores próximos de 1.0. Esse resultado demonstra a robustez do modelo, pois ele alcança uma performance elevada independentemente do subconjunto dos dados utilizado para treino e validação. A consistência entre os *folds* é uma evidência de que o modelo não depende de uma

única divisão dos dados para manter sua acurácia, o que aumenta sua confiabilidade.

Figura 11 – Gráfico de acurácia por *fold* do modelo MLP



Fonte: Autor
Fonte: Autores (2025).

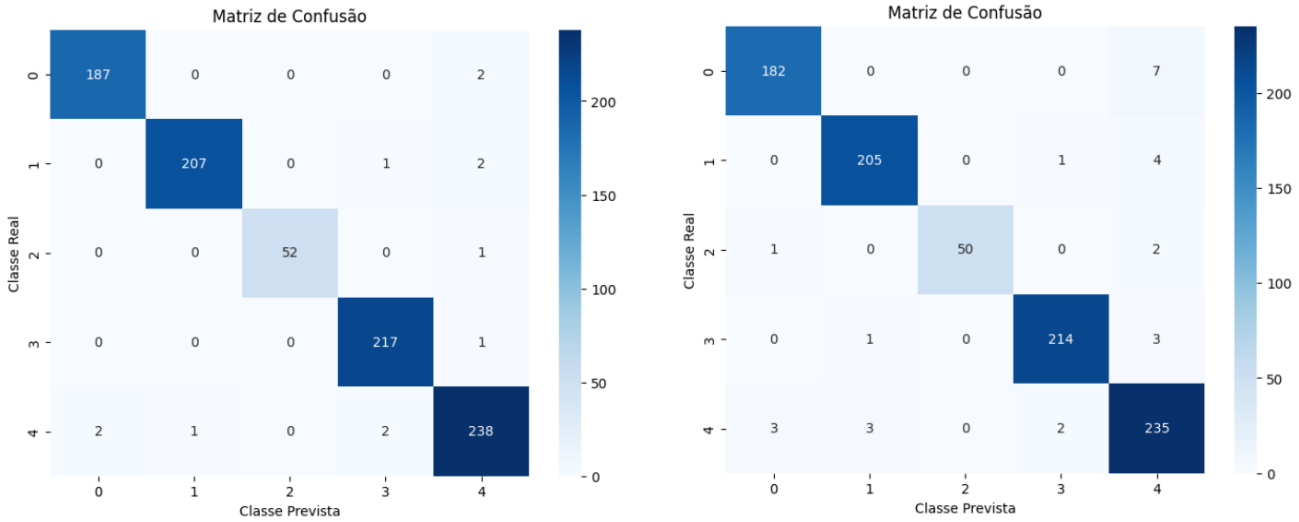
A constância observada nos resultados da validação cruzada reflete a qualidade da abordagem adotada para processar o áudio. A escolha de utilizar 40 coeficientes, em vez dos 13 habituais, permitiu preservar nuances importantes da voz e do timbre de cada magistrado, que seriam perdidas em uma análise mais simplificada. Com essa riqueza de detalhes, o modelo conseguiu distinguir os oradores com maior clareza, mantendo um desempenho uniforme e seguro, independentemente da forma como os dados de treinamento foram divididos.

Esses resultados indicam que o modelo está bem ajustado e é robusto o suficiente para aplicações práticas, minimizando erros de classificação e mantendo estabilidade ao longo de diferentes amostras de validação. A matriz de confusão e os gráficos de acurácia, perda e validação cruzada confirmam o desempenho satisfatório do modelo em distinguir com precisão as classes de áudio.

Análise e desempenho dos demais modelos

Quanto à análise dos resultados dos algoritmos clássicos, observa-se que as matrizes de confusão dos algoritmos SVM, RF, AdaBoost e LR destacam diferenças no desempenho de cada modelo para a classificação de áudios. Na Figura 12, o SVM e o RF se destacaram com alta precisão e poucas confusões, especialmente nas classes mais frequentes. O RF mostrou-se particularmente robusto, com uma baixa taxa de erros em todas as classes, enquanto o SVM teve um desempenho semelhante, embora com algumas confusões em classes específicas, como a Classe 0.

Figura 12 – Comparação entre matrizes de confusão dos modelos usando SVM e RF

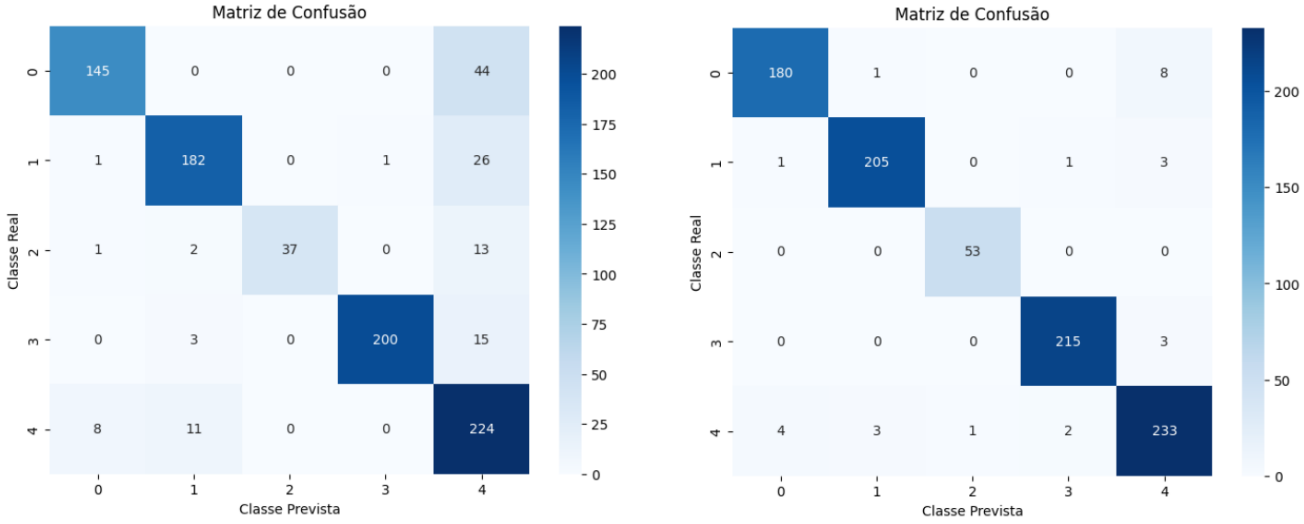


Matriz de confusão modelo com SVM | Matriz de confusão modelo com RF

Fonte: Autores (2025).

Por outro lado, o AdaBoost e a LR apresentaram maior dificuldade na discriminação de certas classes, resultando em um número maior de erros, especialmente nas Classes 0 e 4, conforme observado na Figura 13. O AdaBoost, por exemplo, confundiu um número significativo de instâncias da Classe 0, indicando uma menor capacidade de separar classes próximas. A LR, sendo um modelo linear, também demonstrou limitações em capturar relações mais complexas entre as características dos dados.

Figura 13 – Comparação entre matrizes de confusão dos modelos usando AdaBoost e LR



Matriz de confusão modelo com AdaBoost | Matriz de confusão modelo com LR

Fonte: Autores (2025).

Os resultados obtidos também foram analisados sob a perspectiva de sua transferibilidade para contextos organizacionais semelhantes, considerando seu potencial de uso como referência em consultorias e em atividades formativas baseadas em casos.

Considerações Finais

Este trabalho teve como objetivo desenvolver e avaliar um modelo capaz de identificar automaticamente a fala do magistrado em gravações reais de audiências trabalhistas, com foco em sua aplicação prática no Tribunal Regional do Trabalho da 8ª Região (TRT8). A proposta surge como um complemento à ferramenta de degravação Degrava, já utilizada na instituição, visando aprimorar o processo de análise das sessões e apoiar a elaboração de minutas e votos no Segundo Grau.

Os resultados obtidos demonstram que a abordagem adotada é viável e apresenta desempenho elevado mesmo em um cenário acústico desafiador, caracterizado por ruídos ambientais, variação na qualidade de gravação e sobreposição de falas. Dentre os algoritmos avaliados, o modelo MLP apresentou o melhor desempenho, com acurácia próxima de 99% e métricas consistentes entre todas as classes, incluindo aquelas com menor quantidade de dados. Este desempenho reforça que métodos de aprendizado de máquina de menor complexidade podem ser eficazes mesmo sem o uso de técnicas de aumentação de dados, mantendo-se compatíveis com as limitações de infraestrutura existentes.

Do ponto de vista prático, a solução proposta não elimina, por si só, a morosidade processual, mas contribui de maneira significativa para otimizar o fluxo de trabalho dos analistas e assessores. Ao permitir a identificação automática dos trechos em que o magistrado fala, a ferramenta possibilita localizar com maior rapidez os pontos decisórios das audiências, reduzindo o esforço de navegação manual nos arquivos, mitigando erros de atribuição e oferecendo mais segurança ao processo de transcrição e análise.

Cientificamente, a pesquisa demonstra que modelos tradicionais de classificação, quando combinados a um processo robusto de pré-processamento e extração de características, são capazes de alcançar resultados comparáveis a arquiteturas mais complexas, mesmo trabalhando com dados reais provenientes de ambientes pouco controlados. Institucionalmente, o trabalho contribui para fortalecer uma solução já existente, ampliando sua utilidade e potencial de impacto no TRT8.

Os resultados obtidos indicam que a identificação automática da fala do magistrado é tecnicamente viável, aplicável ao contexto de uso e alinhada às demandas de modernização do Judiciário. A proposta apresenta potencial de integração com ferramentas já adotadas pela instituição e representa um avanço significativo dentro do ecossistema de apoio à análise processual.

Em termos de aplicabilidade, os resultados deste estudo podem servir de referência para processos de consultoria voltados à transformação digital, revisão de fluxos de trabalho e adoção de soluções baseadas em inteligência artificial em instituições públicas e organizações que lidam com tratamento de áudio e análise documental. Além disso, os achados podem ser adaptados para a construção de materiais didáticos e casos de ensino, especialmente em cursos e disciplinas relacionados à inovação, gestão pública, sistemas inteligentes e tecnologia aplicada ao contexto jurídico, contribuindo para aproximar a

produção científica de situações reais de tomada de decisão.

Referências

BERRAR, D. Cross-validation. In: RANGANATHAN, S.; GRIBSKOV, M.; NAKAI, K.; SCHÖNBACH, C.; CANNATARO, M. (ed.). **Encyclopedia of bioinformatics and computational biology**. Amsterdam: Elsevier, 2019. v. 1, p. 542-545.

CHU, S.; NARAYANAN, S.; KUO, C.-C. J. Environmental sound recognition with time–frequency audio features. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 17, n. 6, p. 1142–1158, 2009.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Justiça em números 2024**. Brasília: CNJ, 2024. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2025/04/justica-em-numeros-2024.pdf.

GANCHEV, T.; FAKOTAKIS, N.; KOKKINAKIS, G. Comparative evaluation of various MFCC implementations on the speaker verification task. In: INTERNATIONAL CONFERENCE ON SPEECH AND COMPUTER (SPECOM), 10., 2005. **Proceedings** [...]. [S.l.: s.n.], 2005. v. 1, p. 191–194.

INAZAWA, P. et al. Projeto Victor. **Computação Brasil**, n. 39, p. 19–24, 2019. DOI: https://doi.org/10.5753/compbr.2019.39.

KINGMA, D. P.; BA, J. Adam: a method for stochastic optimization. **arXiv preprint**, arXiv:1412.6980, 2014.

OLIVEIRA, K. W. A. de; PINTO, S. P. da S. O.; TABORDA, E. M. P.; SILVA, R. dos P. S. da; CARVALHO, Z. V. de. Cabos ópticos ADSS em cidades inteligentes: perspectivas para a transformação digital e a transição energética — subsídios para consultoria em infraestrutura e suporte ao ensino de gestão urbana. **Revista de Casos e Consultoria**, Natal, v. 17, n. 1, e42760, 2026. Disponível em: https://periodicos.ufrn.br/casoseconsultoria/article/view/42760. Acesso em: 9 nov. 2026.

SANTOS, J. da S.; AMORIM, I. B. de. Interfaces sociais e saberes em ação: uma análise do estado do conhecimento e suas contribuições para prática em consultorias e ensino baseado em casos. **Revista de**

Casos e Consultoria, Natal, v. 17, n. 1, e43038, 2026. Disponível em: https://periodicos.ufrn.br/casoseconsultoria/article/view/43038. Acesso em: 9 nov. 2026.

SILVA NETO, A. R.; SOUZA, E. F. da S. Aplicação de métodos de previsão de demanda como ferramenta de consultoria organizacional no centro cirúrgico de um hospital privado. **Revista de Casos e Consultoria**, Natal, v. 17, n. 1, e42710, 2026. Disponível em: https://periodicos.ufrn.br/casoseconsultoria/article/view/42710. Acesso em: 9 nov. 2026.

STAATS, J. L.; BOWLER, S.; HISKEY, J. T. Measuring judicial performance in Latin America. **Latin American Politics and Society**, v. 47, n. 4, p. 77–106, 2005. DOI: https://doi.org/10.1111/j.1548-2456.2005.tb00329.x.

TRT8. **Painel de Monitoramento e-Gestão 2º Grau**. 2025. Disponível em: https://app.powerbi.com/view?r=eyJrIjoiZjk1NDBjZjYtOGMxNC00OWY2LTg5MjktNDk4MzQzZmVmYWEzIiwidCI6ImNjZDk5MTdlLW. Acesso em: 16 ago. 2025.