Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Bacharelado em Ciência da Computação

Avaliação automática de pausas de sentido na leitura de estudantes do ensino básico

Cristiano Nascimento da Silva

JUIZ DE FORA
JANEIRO, 2023

Avaliação automática de pausas de sentido na leitura de estudantes do ensino básico

Cristiano Nascimento da Silva

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Jairo Francisco de Souza

JUIZ DE FORA

JANEIRO, 2023

Avaliação automática de pausas de sentido na leitura de estudantes do ensino básico

Cristiano Nascimento da Silva

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Jairo Francisco de Souza
Doutor em Informática, PUC-Rio

Victor Stroele de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação, UFRJ

Heder Soares Bernardino
Doutor em Modelagem Computacional, LNCC/MCTI

JUIZ DE FORA
20 DE JANEIRO, 2023

Ao meu irmão, irmãs, amigos e amigas.

Aos meus pais, pelo apoio e sustento.

Resumo

Com as diversas avaliações formativas de fluência de leitura sendo cada vez mais utilizadas, se faz necessária uma automatização do processo a fim de diminuir os custos, sejam de tempo ou de contratação de pessoal especializado. Por mais que muitos trabalhos automatizem o processo de avaliação da fluência, geralmente avaliam apenas a acurácia e a velocidade na leitura, os quais são dois dos três aspectos que compõem a fluência na leitura. A avaliação desses dois aspectos é o suficiente para avaliar a decodificação das palavras, mas não a compreensão do texto. O terceiro aspecto, a prosódia, é componente facilitador da compreensão leitora. A prosódia é composta por diversos componentes, como entonação, duração, pausa de sentido e acentuação. As pausas de sentido consistem em momentos sem pronúncia durante a fala, responsáveis por auxiliar no significado do que é dito. O presente trabalho tem o objetivo de automatizar a avaliação da pausa de sentido. A avaliação da pausa de sentido pressupõe sua identificação correta durante a fala, o que não é um processo trivial, dado que é necessária a interpretação de sinais de fala de forma automática, e a diferenciação entre pausas que estejam relacionadas ao sentido das que sejam uma decorrência da leitura lenta. Para a avaliação da proposta, foi utilizada uma base de dados composta por áudios de 1 minuto de leitura de textos por crianças nos primeiros anos de escolaridade. Os áudios foram fornecidos pela Fundação CAEd/UFJF, o qual é reconhecido pelo trabalho em avaliação educacional. O presente trabalho contribui para a literatura na medida em que não foram encontrados trabalhos com o foco em avaliar automaticamente o uso de pausa de sentidos na leitura na língua portuguesa. Foi possível identificar de forma automática as ocorrências de pausas utilizando o modelo *Wav2vec2*. Quanto à avaliação automática quanto ao respeito às pausas, foram testados 5 modelos de classificação e 8 *features*, sem verificação de melhores parâmetros para os modelos. O melhor resultado alcançou uma acurácia de 69,5% para uma base desbalanceada de 859 áudios, cujas avaliações manuais de referência podem conter erros.

Palavras-chave: Fluência, prosódia, compreensão, pausa, *Wav2Vec2*, classificação.

Abstract

With the various formative evaluations of reading fluency being increasingly used, it is necessary to automate the process in order to reduce costs, whether in terms of time or hiring specialized personnel. As much as many studies automate the fluency assessment process, they generally only assess reading accuracy and speed, which are two of the three aspects that make up reading fluency. The assessment of these two aspects is enough to assess word decoding, but not text comprehension. The third aspect, prosody, is a component that facilitates reading comprehension. Prosody is composed of several components, such as intonation, duration, pause of meaning and stress. Sense pauses consist of moments without pronunciation during speech, responsible for helping in the meaning of what is said. The present work aims to automate the evaluation of the pause in meaning. The evaluation of the pause in meaning considered its correct identification during speech, which is not a trivial process, given that it is necessary to interpret speech signals automatically, and the differentiation between pauses that are related to the meaning of those that are a as a result of slow reading. For the evaluation of the proposal, a database composed of 1-minute audios of text reading by children in the first years of schooling was used. The audios were provided by Fundação CAEd/UFJF, which is recognized for its work in educational evaluation. The present work contributes to the literature insofar as no works were found that focused on automatically evaluating the use of pauses in meaning in reading in Portuguese. It was possible to automatically identify occurrences of pauses using the *Wav2vec2* model. As for the automatic evaluation regarding respect for pauses, 5 classification models and 8 *features* were tested, without checking the best parameters for the models. The best result achieved an accuracy of 69.5% for an unbalanced base of 859 audios, whose manual reference evaluations may contain errors.

Keywords: Fluency, prosody, comprehension, pause, *Wav2Vec2*, classification.

Agradecimentos

A todos os meus parentes, pelo encorajamento e apoio.

Ao professor Jairo pela orientação, amizade e principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

"Lembra que o sono é sagrado e alimenta de horizontes o tempo acordado de viver".

Beto Guedes (Amor de Índio)

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Abreviações

DCC Departamento de Ciência da Computução

UFJF Universidade Federal de Juiz de Fora

1 Introdução

As avaliações formativas com testes de fluência em leitura têm sido cada vez mais utilizadas em estados e municípios brasileiros (SOUSA, 2014), porém, estes se mostram custosos, pois demandam a contratação de avaliadores humanos treinados (CARCHEDI; BARRÉRE; SOUZA, 2021). Essas avaliações podem ser utilizadas para tomadas de decisão quanto a novas abordagens de ensino de acordo com o desempenho dos(as) alunos(as), como Sousa (2014) fala sobre seu uso como "referência para iniciativas de gestão".

Uma automatização de pelo menos parte do processo de avaliação se faz necessária para que se possa reduzir os custos que a avaliação demanda, e o tempo necessário para se obter resultados (CARCHEDI; BARRÉRE; SOUZA, 2021). A automatização pode ser feita em relação a diversos aspectos da fluência, pois as avaliações de fluência são compostas geralmente de três aspectos: acurácia, velocidade e prosódia (BOLAÑOS et al., 2013a; BOLAÑOS et al., 2013b; BENJAMIN et al., 2013).

1.1 Apresentação do Tema

Tendo a prosódia como componente da fluência, ela tem sua importância na avaliação completa da fluência. Além disso, um dos componentes que fazem parte do conceito de prosódia, é a pausa de sentido, foco do presente trabalho.

De acordo com Godde, Bailly e Bosse (2019), muitos trabalhos que visam a criação de sistemas computacionais avaliadores de leitura tendem a focar apenas nos componentes de acurácia e velocidade que compõem a fluência, contando o número de palavras lidas pelo leitor, bem como a velocidade com que a leitura é realizada. Porém, avaliações feitas com foco em apenas esses aspectos conseguem identificar a capacidade de se decodificar palavras, mas não de entender seus significados (GODDE; BAILLY; BOSSE, 2019). Isso causa uma sub-representação da fluência em leitura, de acordo com Bolaños et al. (2013a). Assim, percebe-se a importância da avaliação da prosódia na avaliação da fluência em leitura, e assim, mostrando também a importância da avaliação das pausas de sentido,

dado que compõem a prosódia.

1.2 Descrição do Problema

A avaliação de prosódia traz diversas dificuldades pois é um conceito que integra diversos componentes, cujos nomes variam nos trabalhos que os citam (PULIEZI; MALUF, 2014; VEENENDAAL; GROEN; VERHOEVEN, 2015; BARBOSA, 2012), mas os conceitos se interceptam. De acordo com Puliezi e Maluf (2014), os componentes da prosódia são a entonação, acentuação tônica, a duração e a pausa.

A detecção automática da prosódia não é uma tarefa trivial, pois a representação de eventos prosódicos de forma categórica possui ambiguidades (SRIDHAR; BANGALORE; NARAYANAN, 2008). Além disso, para avaliar a prosódia automaticamente, é necessário que se tenha um modelo acústico adequado, como Yang et al. (2022) mostrou em seu trabalho. Isso se deve ao fato de que os componentes da prosódia dependem de características acústicas da fala, então é necessário que o modelo seja capaz de coletar as características acústicas do áudio de forma correta. Isto que requer uma base de dados suficientemente grande ou suficientemente específica para que o modelo de reconhecimento automático consiga identificar o que foi dito de forma adequada.

Portanto, algumas questões são verificadas com o presente trabalho. Primeiro se é possível identificar as pausas de sentido na fala de forma automática e confiável. Isso é importante pois para avaliar as pausas, primeiro deve-se identificá-las adequadamente. Então ao ter essas informações das pausas, é investigado o impacto delas para a avaliação de fluência na leitura. E por fim, é verificado se é possível avaliar automaticamente o respeito às pausas com base nessas informações.

1.3 Contextualização

Dos componentes da prosódia, a pausa se refere ao momento da fala em que não há fonação e pode ser classificada em pausa por falta de ar ou por fator significativo (PULIEZI; MALUF, 2014). Existem trabalhos que mostram a importância da prosódia para a análise da fluência, pois mostram que a prosódia tem impacto na compreensão da leitura

(BARROS, 2017; LOPES et al., 2015). A prosódia pode ser considerada como um indicador de maior compreensão na leitura, bem como uma catalisadora disso (BENJAMIN; SCHWANENFLUGEL, 2010). Na medida em que a leitura da criança evolui ao ponto de não haver preocupação com a decodificação de palavras, há maior atenção dela para a prosódia e compreensão do texto a ser lido (BOLAÑOS et al., 2013a; SCHWANENFLUGEL; BENJAMIN, 2017; GODDE; BAILLY; BOSSE, 2019).

Com isso, como explicado em (BENJAMIN et al., 2013), a prosódia permite que informações importantes do texto sejam guardadas na memória durante a leitura, frases lidas com mais expressividade são lembradas com facilidade por adultos e crianças. A prosódia indica quais partes de uma redação merecem mais atenção e ainda as medidas de expressividade de crianças nas primeiras séries escolares ajudam a prever a compreensão de texto das mesmas em anos posteriores (BOLAÑOS et al., 2013a; SCHWANENFLUGEL; BENJAMIN, 2017).

1.4 Justificativa/Motivação

Dada a prosódia como componente que compõe a fluência em leitura (FERREIRA, 2009), o presente trabalho se justifica na medida em que para um sistema que avalia fluência em leitura esteja completo, se faz necessária a avaliação da prosódia na leitura. Logo, identificar corretamente os componentes da prosódia, é importante para que todos os aspectos da fluência sejam observados. O presente trabalho tem como objetivo identificar e avaliar o uso de pausas de sentido em áudios de leituras de estudantes do ensino básico.

Dessa forma, o presente trabalho deve contribuir para uma complementação da avaliação de leitura. Deve contribuir também como um estudo de abordagens para a identificação automática de pausas de sentido, dado que no processo de descobrir a melhor abordagem, diferentes abordagens foram testadas.

1.5 Objetivos

O trabalho tem como objetivo a identificação automática de pausas de sentido (componente da prosódia) presentes em áudios de leituras textuais feitas por crianças. A

identificação dessas pausas tem objetivos práticos e teóricos, dentre eles:

- Contribuir para uma melhor avaliação automática da qualidade de leitura de textos, dado que a prosódia é componente facilitador da compreensão (BARROS, 2017).
- Investigar componentes acústicos presentes na fala que auxiliem na identificação automática de pausas de sentido.
- Possibilitar a criação de um algoritmo preciso utilizando um modelo acústico treinado para português, para a identificação de pausas de sentido durante a fala.

1.6 Metodologia

Para a identificação adequada das pausas de sentido em áudios de crianças, se faz necessário um modelo acústico bem treinado para conseguir reconhecer corretamente os fones pronunciados. Dessa forma, o modelo (*Wav2vec2*) utilizado foi treinado com uma base suficientemente grande em leituras em português e se necessário, com acréscimo de áudios de crianças, para ser mais específico, dado que as leituras das crianças tendem a ter mais pausas e serem menos fluidas que as dos adultos.

O método pensado consiste em utilizar de análises espectográficas utilizando o *Wav2Vec2*, que é um modelo acústico pré-treinado para reconhecimento automático de fala, o qual é utilizado por Yang et al. (2022) para auxiliar no reconhecimento de pronúncias erradas.

A base disponível para teste consiste de 50 mil áudios de leituras textuais de crianças, fornecidas pela Fundação CAEd/UFJF, mas foi utilizado um subconjunto menor dessa base, composto de áudios que contenham as informações de classificação manual quanto ao respeito às pausas para a comparação com resultado automático. O subconjunto também só pode ter áudios que tenham sido reconhecidos automaticamente pelo *Wav2vec2*, de forma a ser possível a extração das pausas. Os áudios têm uma duração média de 1 minuto e fazem parte das leituras de um dos itens avaliativos de uma de suas avaliações formativas já realizadas em larga escala. Foi concedido também acesso ao resultado da avaliação manual de cada um dos áudios. Dentre os diversos dados sobre

a leitura de cada áudio, há uma classificação binária sobre o áudio como um todo, para dizer se a criança obedeceu ou não às pausas de sentido durante a leitura.

O objetivo é criar um algoritmo que utiliza as informações acústicas fornecidas pelo modelo, para classificar o áudio de forma binária, em presença ou ausência de utilização correta de pausas na maior parte do tempo do áudio.

O *Wav2Vec2* foi utilizado para o processamento dos áudios, de forma que foi possível obter uma matriz de probabilidades para cada letra das palavras esperadas de serem lidas e o que foi de fato pronunciado no áudio em cada momento do tempo.

Foi utilizada então as informações acústicas de cada *frame* do áudio, para alinhar o que foi dito com o texto esperado e identificar o início e o final de cada palavra lida, de forma a identificar assim o silêncio entre cada leitura. Dessa forma, foi possível uma tomada de decisão sobre o áudio todo com base em todas as pausas reconhecidas automaticamente.

2 Fundamentação Teórica

Para entender melhor a necessidade da identificação da pausa, como componente da prosódia, no contexto de avaliação da fluência, o presente capítulo trata de uma fundamentação teórica acerca do tema. Este capítulo está dividido de forma que a seção 2.1 traz definições do termo prosódia de acordo com diferentes trabalhos, a seção 2.2 fala sobre trabalhos que estudam a relação da prosódia com a alfabetização. Já a seção 2.3, explica os componentes que fazem parte da prosódia, enquanto a seção 2.4 estuda utilizações da prosódia na educação básica. à utilização de Inteligência Artificial na avaliação de prosódia, passando pela sua importância na alfabetização e na fluência em leitura.

2.1 Definição de prosódia

Um conceito importante a ser compreendido para este trabalho é o de "prosódia". O significado de prosódia, de acordo com o Dicionário Online de Português é "Parte da gramática normativa que trata da reta acentuação dos vocábulos e, ainda, dos fenômenos de entoação" (PROSóDIA, 2018), ou seja, prosódia tem relação com as propriedades acústicas da fala, indicando que apenas o aspecto ortográfico não se faz suficiente para a identificar.

No trabalho de Barbosa (2012), a prosódia é apresentada nos fatores linguísticos, paralinguísticos e extralinguísticos, mas, para o presente trabalho, o importante é o aspecto linguístico do conceito. Na Linguística então, de acordo com Barbosa (2012), a prosódia está relacionada ao "modo de falar" do que está sendo dito. É possível perceber que o conceito apresentado carrega uma subjetividade, o que dificulta sua identificação automática, pois é necessária uma definição mais concreta para que seja possível saber quais características da fala devem ser observadas na identificação da prosódia.

Já de acordo com BARROS (2017), a prosódia "se caracteriza pelo uso apropriado do fraseado e da expressividade para transmitir um significado ao longo de um texto". Essa definição demonstra que a prosódia carrega uma intenção na fala, auxiliando na

transmissão de seu significado. A prosódia é um conceito muito importante na medida em que está diretamente relacionada ao significado da frase pronunciada.

Em (PULIEZI; MALUF, 2014), por sua vez, é considerado que ler com prosódia, é o mesmo que ler com expressão adequada, com ritmo e entonação, de forma a possibilitar a manutenção do significado. Puliezi e Maluf (2014) atrelam a prosódia a uma leitura boa e não simplesmente a características da leitura. As autoras definem também a prosódia como sendo a "música da linguagem oral". A definição traz a ideia de que a prosódia é composta por diferentes aspectos e também que está fortemente ligada às características acústicas da fala.

As autoras em (VEENENDAAL; GROEN; VERHOEVEN, 2015) dizem que ler com prosódia, faz com que a leitura em voz alta soe mais natural a partir do fraseamento adequado, a utilização de pausas, limites de palavras e frases e expressividade no geral. A definição apresentada por Veenendaal, Groen e Verhoeven (2015) é carregada de subjetividade, assim como em (BARBOSA, 2012), e da integração de diversos componentes.

Por mais que as definições variem entre autores, todos os trabalhos citados continuam apontando a prosódia como um conceito relacionado a uma leitura mais adequada, que depende da interpretação de informações presentes no sinal de fala e que é composta por diversos componentes. Veenendaal, Groen e Verhoeven (2015) mostram que a prosódia é composta por diversos componentes quando dizem que depende do fraseamento, pausas, limites de palavras e frases, e expressividade no geral, por mais que expressividade seja um conceito sinônimo a prosódia em trabalhos como (BOLAÑOS et al., 2013a).

Alguns trabalhos ainda apresentaram uma relação entre fluência na leitura e prosódia (VEENENDAAL; GROEN; VERHOEVEN, 2015; BARROS, 2017; PULIEZI; MALUF, 2014). Todos esses trabalhos, colocam a prosódia como componente da fluência, além de fator relacionado à compreensão. Isso reforça a importância de se avaliar prosódia no contexto de avaliação de fluência na leitura.

2.2 Prosódia na alfabetização

Quanto à relação da prosódia com a alfabetização, Pardinho (2021) fala que crianças em idade de 7 a 11 anos compreendem melhor o texto quando este traz variação prosódica.

Relaciona ainda alfabetizar a uma leitura que produz sentido, dizendo que a única maneira de um texto produzir sentido, é se ele for lido com variação prosódica adequada ao contexto. Dessa forma, a autora caracteriza a prosódia como elemento fundamental no processo de alfabetização.

Em (BARROS, 1994), também é afirmado que a prosódia tem influência significativa no sentido de um texto produzido oralmente. Afirma ainda que a prosódia é como uma "chave de interpretação" que o falante fornece ao interlocutor, de forma a guiá-lo. Quanto à alfabetização, a autora considera a prosódia como apropriada para aquisição da linguagem oral e escrita adequada, pois a prosódia tem também função de conexão de elementos do texto, auxiliando na coesão textual, de acordo com Barros (1994).

2.3 Componentes que integram na prosódia

Conforme falado anteriormente, vários trabalhos dividem a prosódia em componentes, alguns até mesmo relacionam sua definição ao bom emprego de seus componentes. Em (BARBOSA, 2012), é afirmado que a prosódia está ligada a diferentes elementos linguísticos como "acento, fronteira de constituinte, ênfase, entoação e ritmo".

Quanto aos componentes, Barbosa (2012) relaciona a "fronteira de constituinte" ou "limite de constituinte", à função de demarcação da prosódia, em "sílabas, palavras fonológicas e grupos acentuais". Diz ainda que a pausa durante a fala indica essas fronteiras de constituinte.

Quanto à entoação e ritmo, o autor diz que em uma perspectiva em que são independentes entre si, a entoação "restringe-se à análise, ao longo do enunciado, das variações de altura, ou seja, das sensações de grave e agudo". Enquanto o ritmo "compreende as variações de duração percebida de unidades do tamanho da sílaba ao longo do enunciado". Ou seja, a entoação está atrelada a variações de altura da fala e o ritmo às de duração.

Com relação ao acento frasal, ele diz que é "a indicação da proeminência de um elemento do enunciado, realizado em torno da sílaba tônica ou lexicalmente acentuada". E ênfase ele diz ser a "proeminência manifesta de uma unidade linguística com função de insistência ou para chamar a atenção para uma informação crucial, entre outras funções".

Em (VEENENDAAL; GROEN; VERHOEVEN, 2015), a prosódia é relacionada

à boa utilização de fraseamento, pausas, limites de palavras, limites de frases e expressividade. Mas quando são citadas as características prosódicas presentes na leitura de um texto, as autoras falam em "pausas e frequência fundamental".

Quanto a essas características, a frequência fundamental citada, remete a "uma das medidas mais importantes da análise acústica e tem relação direta com o comprimento, tensão, rigidez e massa das pregas vocais e estas com a pressão subglótica", de acordo com Braga, Oliveira e Sampaio (2009). Ou seja, é uma característica relacionada ao cumprimento de onda da fala.

Já em (PULIEZI; MALUF, 2014), a prosódia é separada em 4 componentes, definindo cada um deles. Os componentes são a entonação, a acentuação tônica, duração e a pausa.

As autoras definem a entonação como sendo a "frequência da fala". De acordo com elas, a entonação indica se a frase está perto de uma pausa ou de seu fim ou se ainda continuará. Se a entonação for crescente é porque a frase continuará, já a decrescente indica o contrário.

Quanto à acentuação tônica, é a característica que remete à tonicidade da palavra. A duração é caracterizada pelas autoras como o "tempo de articulação de um som, sílaba ou enunciado e tem uma importância fundamental no ritmo de cada língua". Quanto mais rápida na leitura uma pessoa é, menor será a duração da pronúncia das palavras.

A pausa é definida em (PULIEZI; MALUF, 2014) como sendo "uma unidade de tempo onde não há fonação", isto é, um período de tempo em que não há pronúncia alguma. As autoras explicam que as pausas podem ser originadas pela falta de ar nos pulmões ou pelo fator significativo, que é o mais relevante para a prosódia. É dito também que como as frases são unidades de sentido, elas são compostas por pequenas unidades de sentido, que são as palavras, as quais são delimitadas por pausas.

Por mais que utilizem nomes diferentes para os componentes da prosódia, os significados de seus componentes se interceptam. O que alguns autores chamam de "ritmo", outros chamam de "duração". O que alguns chamam de "entonação", outros chamam de "frequência fundamental". São nomes diferentes, mas remetem a um mesmo significado.

Como a prosódia abrange vários componentes, o presente trabalho tem como

objetivo identificar de forma automática a presença de um dos aspectos que envolvem a prosódia, a pausa de sentido, também chamada pelo Barbosa (2012) de "fronteira de constituinte".

2.4 Investigações de prosódia na educação básica

Prosódia pode ser utilizado como um dos componentes avaliativos para avaliação de leitura em voz alta. Há diversas formas de avaliar a prosódia durante a leitura, através da análise de características da leitura, como tom, expressão, separação das frases. Há também os que avaliam a compreensão textual e relacionam com aspectos da leitura, como distinção de palavras simples e compostas de acordo com o sentido.

(FERREIRA, 2009) é um exemplo de trabalho que analisa tom e outras características da fala. Ferreira (2009) avalia diversos aspectos na leitura com o objetivo de avaliar a fluência em leitura em 142 alunos no final do 2º ano de escolaridade. Um dos aspectos avaliados é a prosódia. Para isso, foram realizados dois estudos, o primeiro para avaliar questões "psicométricas (índice de dificuldade, poder discriminativo, fidelidade, validade externa) e o segundo com finalidade de comparar a fluência de leitura dos alunos com e sem Necessidades Educativas Especiais". No segundo estudo, foi utilizado o texto com as melhores qualidades psicométricas e depois foi feita a separação dos resultados dos com e sem NEE (Necessidades Educativas Especiais) em aspectos da fluência, que são a precisão, prosódia e velocidade.

Para avaliar a prosódia, avaliadores ouviram a leitura em voz alta de cada criança e fizeram observações. Segundo o autor, durante a leitura o avaliador foi instruído a ouvir o tom, a expressão e a separação de partes das frases. Alguns aspectos observados durante a leitura foram a ênfase vogal nas palavras necessárias, se a entoação bate com a pontuação textual e a utilização de diferentes características na leitura, como pontuações e conjunções, para verificar se ocorreram pausas de maneira adequada. Foi verificado então por Ferreira (2009) que o método de análise da prosódia fez com que fosse possível distinguir o desempenho dos estudantes.

Já em (FILIPE; VICENTE, 2010), a capacidade de segmentação prosódica é utilizada para avaliar a compreensão e produção sintática de frases ambíguas, que basicamente

consiste na capacidade de distinção de palavras simples e compostas e de desambiguação de palavras. Os teste foram feitos em 43 crianças do 1º ciclo do ensino básico, 10 adultos e 12 crianças com síndrome de Asperger. Para o teste, foi utilizada a "prova de Segmentação do Profiling Elements of Prosodic Systems-Children", de sigla PEPS-C, "adaptada para o Português Europeu". A prova é dividida em duas tarefas, a primeira avalia a capacidade de distinguir entre palavras simples e palavras compostas, ao fazer o aplicante ver imagens e pronunciar as palavras que as representam. Como exemplo de palavras a serem faladas para representar as imagens apresentadas, Filipe e Vicente (2010) citam "porta, chaves e leite" e "porta-chaves e leite". A segunda tarefa verifica as capacidades de desambiguação de palavras pelo sentido ao utilizar imagens com meias de várias cores (FILIPE; VICENTE, 2010). Essa tarefa solicita que se escolha imagens de meias coloridas que corresponda à frase ouvida. A autora utiliza como exemplo as frases "meias pretas&verdes e rosas" e "meias pretas e verdes&rosas". Com esses testes, foi possível analisar a relação entre idade e os resultados obtidos nos dois testes. Os resultados mostraram "ganhos desenvolvimentais na competência de segmentação prosódica em função da idade" (FILIPE; VICENTE, 2010), pois a competência prosódica de segmentação não está completamente desenvolvida até os 11 anos de idade, mas os adultos a utilizam de forma eficaz (FILIPE; VICENTE, 2010).

Em (PINTO; NAVAS, 2011), foi verificada a influência da leitura por meio de programa de estimulação de leitura, baseado em padrões de prosódia, como variação de entonação, duração e aceleração, no desempenho da fluência de leitura de 32 crianças no quinto ano do Ensino Fundamental. O teste consistiu em leituras de textos e descrição de uma imagem. O teste avaliou diversos aspectos na leitura, como: taxa de leitura, velocidade de fala, compreensão de textos e adequação da variação da prosódia durante a leitura. Após a aplicação do teste, as crianças participaram de um programa, no sentido de planejamento, que estimula a leitura com ênfase na prosódia. Esse programa é composto por 5 sessões de 15 minutos cada, realizadas uma vez por semana. Após isso, o teste foi realizado novamente para se verificar qual o desempenho obtido após o estímulo do programa. Os resultados mostraram que o desempenho melhorou após a utilização do programa, indicando a prosódia como elemento influenciador do desempenho em leitura

como um todo.

Os resultados observados nos trabalhos apresentados (FILIPE; VICENTE, 2010; FERREIRA, 2009; PINTO; NAVAS, 2011) mostram que a prosódia é um bom parâmetro de análise de desempenho na leitura, que é parte importante do aprendizado e da melhoria da leitura e que crianças em idades iniciais possuem dificuldade maior em sua utilização. Tendo em vista que avaliações formativas podem ser utilizadas para tomadas de decisão quanto a planejamentos de ensino, esses resultados reforçam a importância da avaliação da prosódia em testes formativos principalmente nos primeiros anos de escolaridade.

2.5 Avaliação de prosódia

Há também trabalhos em que o foco da avaliação é a própria prosódia, como é o caso de BARROS (2017), cuja definição de prosódia já foi citada anteriormente. Esse trabalho estuda a relação entre compreensão leitora e prosódia em crianças. Nela, é definida de forma resumida a compreensão leitora como "construção de significados, por meio de geração de inferências, ao articular informações explícitas no texto com o conhecimento de mundo do leitor".

Foram realizados dois estudos, o primeiro em 84 crianças do 3º ano do ensino fundamental e com 40 alunos do 5º ano do ensino fundamental. Os alunos do 3º ano tinham idades de 8 e 9 anos e os do 5º tinham idade de 10 e 11 anos. O segundo estudo foi em 47 dos 84 alunos do 3º ano do ensino fundamental que participaram do primeiro estudo.

O primeiro estudo tinha o objetivo de investigar a relação entre compreensão leitora e prosódia. Para realizar a análise de prosódia e compreensão, foi solicitada a leitura em voz alta de um texto, e para a compreensão, foram realizadas perguntas acerca do texto. A análise da prosódia foi em relação à expressão e volume, fraseado, suavidade e ritmo. O segundo estudo teve foco maior na compreensão leitora e testou situações de leitura (oral e silenciosa) e tipos de discurso (direto e indireto). Esses testes foram feitos de forma a verificar se uma situação ou outra iria causar um maior uso de recursos prosódicos, podendo favorecer a compreensão leitora.

Dos seus estudos realizados em (BARROS, 2017), o primeiro mostrou que a

relação entre os dois conceitos se altera com o passar dos anos escolares, e o outro mostrou que a prosódia é uma facilitadora da compreensão. A principal conclusão desse trabalho é a de que "o efeito facilitador da prosódia para a compreensão só pode ocorrer quando uma criança já domina a decodificação e possui um nível adequado de compreensão leitora".

Esse resultado mostra a importância da prosódia como método avaliativo de fluência em leitura, dada a sua relação com a compreensão leitora.

O trabalho (LOPES et al., 2015) verificou a evolução da compreensão da leitura e da prosódia em 98 crianças entre o 2º e 3º ano de escolaridade. A evolução foi verificada em 4 momentos do tempo. Para avaliar a prosódia, foi utilizado um instrumento que avalia a prosódia na leitura em voz alta, o "Multidimensional Fluency Scoring Guide" (MFSG), que é uma escala que pode fornecer informações formativas para orientar a instrução, bem como informações sumativas e foi apresentada em (RASINSKI, 2004). As escalas são utilizadas para avaliar a fluência do leitor em 4 dimensões: expressão e volume, fraseado, suavidade e ritmo. Segundo Rasinski (2004), as pontuações vão de 4 a 16 e, geralmente, quando a pontuação está abaixo de 8, isso indica que a fluência pode ser uma preocupação. Pontuações de 8 ou acima indicam uma boa progressão na fluência. Em (LOPES et al., 2015), foi observado que "todos os aspectos da prosódia evoluem gradualmente do 2º para o 3º ano de escolaridade". Além disso, foi notado que dos 4 momentos de análise de desempenho, a evolução do primeiro para o segundo momento, foi a menor. Os autores acreditam que isso se deve ao fato das crianças serem leitores iniciantes, a maior parte dos seus esforços mentais eram alocados à decodificação.

Os autores concluíram então que a prosódia deve ser tratada após o nível necessário de decodificação ser atingido. Ainda, Lopes et al. (2015) concluíram que "os resultados mostram que resultados elevados na prosódia estão associados a melhores resultados na compreensão da leitura em qualquer dos momentos de avaliação". Todos os trabalhos que verificaram a influência da prosódia na fluência ou na compreensão da leitura observaram que ela tem um grande impacto no desempenho da qualidade de leitura.

Não foram encontrados trabalhos que avaliam a utilização de pausas de sentido durante a leitura, mas os trabalhos apresentados (BARROS, 2017; LOPES et al., 2015) utilizaram da pausa como parte da avaliação da prosódia, mostrando sua relevância. As

relações das pausas com a fluência em leitura são investigadas em (LOPES et al., 2015), relacionando pausas mais longas a leitores menos fluentes. As pausas são utilizadas por BARROS (2017) como característica da suavidade da leitura, utilizada na avaliação da prosódia. Lopes et al. (2015) também utilizam a pausa na escala de suavidade que o MFSG analisa. O presente trabalho pretende avaliar especificamente as pausas de sentido, de forma automática, dada a sua relação com a fluência em leitura, como componente da prosódia.

2.5.1 Uso de IA na avaliação de prosódia

Dentro do contexto de processamento automático de fala, a prosódia é muito utilizada para auxiliar no processo de síntese de fala, que consiste em transformar um texto em uma fala. Há diversos trabalhos sobre sua utilização nesse sentido (JR et al., 2004; SKERRY-RYAN et al., 2018; ZHANG; SONG; SONG, 2007). A prosódia tem sido utilizada como tentativa de deixar a fala gerada a mais próxima da natural possível. O presente trabalho se propõe a trabalhar em um contexto inverso à síntese de fala, que é o reconhecimento automático de fala (ASR). Dentro desse contexto, foram encontrados poucos trabalhos, como (SABU et al., 2017; SHAHNAWAZUDDIN et al., 2020; ANANTHAKRISHNAN; NARAYANAN, 2009). A utilização de prosódia em sistemas ASR com a finalidade de avaliação da leitura é mais raro ainda. Shahnawazuddin et al. (2020), Ananthakrishnan e Narayanan (2009) usam prosódia para aumentar o sistema ASR como meio para deixar o sistema de reconhecimento de fala mais eficiente, mas apenas em (SABU et al., 2017) ela é utilizada para avaliação de fluência.

Em (SABU et al., 2017), é feita uma avaliação automática de fluência e acurácia em segunda língua de crianças. Para o reconhecimento de fala, os autores utilizam um modelo de linguagem flexível, juntamente a um modelo acústico baseado em uma rede neural profunda, chamado de modo Tandem. Nesse modo, a rede neural funciona como um extrator não linear de características do áudio que alimenta um modelo GMM-HMM, que é um modelo Gaussiano em conjunto com um modelo oculto de Markov. Essas características são combinadas a características MFCC, convencionalmente usadas em ASR.

A pontuação sobre prosódia é dada sobre estimativa de fraseado e predição de proeminência. Para a estimativa de fraseado, primeiro procurando leituras na forma de lista, em que a leitura é pausada. Para isso, é observada na sentença o número de pausas, desvio padrão nas durações das pausas e duração média da sílaba, enquanto que para palavras que não sejam palavras finais de frase, é observada a duração média da sílaba e o intervalo de altura. Caso não seja uma leitura em forma de lista, é procurada a posição em que a fala recomeça ou pausa é observada. A classificação é dividida de forma que a classificação 3 é dada para os casos em que todas as pausas correspondem às esperadas, e se o número de pausas for maior ou menor que o número esperado ou se a pausa estiver na posição errada, é dada a classificação 2. Para a predição de proeminência, é treinada uma árvore de decisão, e se qualquer palavra em um enunciado for predita proeminente, é classificado como 2 o enunciado. Se nenhuma palavra for encontrada proeminente em todo o enunciado, o enunciado é marcado como classificação 1. Esse método obteve resultados razoáveis de acordo com os autores.

Mesmo não havendo muitos trabalhos que utilizem prosódia no contexto de ASR, os trabalhos apresentados mostraram como a prosódia influencia positivamente na eficiência de um sistema ASR (ANANTHAKRISHNAN; NARAYANAN, 2009; SHAH-NAWAZUDDIN et al., 2020; SABU et al., 2017).

2.6 Modelos de classificação

Para o presente trabalho, alguns modelos de classificação são treinados e utilizados, os quais são explicados na presente seção. Foram utilizados 5 modelos, sendo eles o *Random Forest Classifier*, *Logistic Regression*, *Voting Classifier*, *XGBoost* e *MLP Classifier*.

O *Random Forest Classifier*, de acordo com Pal (2005), consiste na utilização de diversos classificadores de árvore de forma que cada classificador é gerado usando um vetor aleatório amostrado independentemente da entrada. Cada árvore gerada então contribui com um voto unitário para a classe mais popular a fim de classificar o vetor de entrada.

Quanto ao *Logistic Regression*, como o próprio nome diz, é um tipo de regressão, então é importante que se entenda o que é uma regressão primeiro. Em (CONNELLY, 2020), regressão linear é definida como um procedimento em que utiliza-se do valor de

uma variável para encontrar outra, de forma que se espere uma distribuição normal da variável dependente e uma relação linear entre as duas. A regressão logística então utiliza variáveis independentes que podem ser nominais, ordinais, intervalos e até razões para predizer os valores de variáveis dependentes dicotômicas.

O *Voting Classifier* simplesmente reúne diferentes predições de classificadores para escolher a classe mais votada entre eles (EL-KENAWY et al., 2020). Neste caso os classificadores utilizados foram o RFC, LR e o algoritmo de classificação chamado *Gaussian Naive Bayes*. Segundo Kamel, Abdulah e Al-Tuwaijari (2019), *Naive Bayes Classifier* é um classificador probabilístico simples, o qual utiliza da aplicação do teorema de Bayes, considerando cada variável de atributo como variável independente. O diferencial do modelo utilizado é que ele considera que os dados estão seguindo uma distribuição Gaussiana.

Conforme explicado em (QIU et al., 2022), "*Extreme gradient boosting* (XGBoost) é um algoritmo baseado em uma árvore de aumento de gradiente, que pode desempenhar um papel poderoso no aprimoramento de gradiente". Como o XGBoost é baseado na teoria da árvore de classificação e regressão, ele tende a ser um método eficaz para resolver problemas de regressão e classificação.

O MLP (*Multi-layer Perceptron*) é um modelo de rede neural multi-camada do tipo *FeedFoward*. O algoritmo usado pelo MLP possui duas fases, conforme abordado em (WINDEATT, 2008). A primeira sendo uma simulação direta para o padrão de treinamento atual, permitindo o cálculo do erro. A segunda calcula para cada peso na rede, como uma pequena alteração afetará a função de erro, de forma que o algoritmo tenta chegar nos melhores pesos para classificar corretamente os dados.

2.7 Considerações Finais

Neste capítulo foram apresentadas definições acerca da prosódia, sua importância como componente da fluência e seus componentes. Através dessas caracterizações, foi mostrada a relevância de identificar e avaliar automaticamente um dos componentes que compõem a prosódia, a pausa de sentido. O capítulo trouxe a utilização da prosódia no contexto de avaliação da leitura e na alfabetização, reiterando a influência da prosódia para uma

avaliação mais completa da fluência em leitura. A utilização de Inteligência Artificial para avaliação de prosódia se mostrou rara, o que contribui para a relevância do trabalho, na medida em que não foram encontrados muitos trabalhos que utilizam de ASR para avaliação da prosódia. O presente trabalho tem como objetivo encontrar a melhor abordagem para identificar automaticamente as pausas de sentido na leitura de crianças em fase de alfabetização, e utilizá-la como parte da avaliação automática da fluência em leitura.

3 Trabalhos relacionados

Há diversas formas de analisar automaticamente a prosódia, e na literatura existem trabalhos que analisam essas possibilidades, seja com o objetivo de identificar características da leitura em voz alta, seja com o objetivo de avaliar a fluência em leitura e seus aspectos. De acordo com Bolaños et al. (2013a), existem duas formas bem estabelecidas de avaliação de prosódia, também chamada de expressividade pelos autores.

A primeira forma é através de métricas, como a do NAEP, *National Assessment of Educational Progress*, que é uma avaliação nacional dos Estados Unidos, organizada pelo *National Center for Education Statistics* (NCES), que calcula o conhecimento de estudantes em diversas áreas do conhecimento. O *National Assessment of Educational Progress of Oral Reading Fluency* (NAEP ORF) (BENJAMIN et al., 2013; BERGNER; DAVIER, 2019) é a avaliação de leitura do NAEP. Ela avalia a fluência na leitura oral, reconhecimento de palavras e decodificação fonológica. Destas, a fluência na leitura oral é a diretriz que mede, entre outras coisas, a expressividade.

A segunda forma bem estabelecida de analisar prosódia na leitura é por meio de análises espectrográficas. Na prática, isso se dá por meio do entendimento de como algumas medidas prosódicas, como variação de frequência, tom, ênfase e pausas no áudio examinado (BOLAÑOS et al., 2013a), se desenvolvem ao longo da leitura de uma pessoa. Essas propriedades são obtidas por meio de *features* (em português, características) coletadas dos áudios.

Nas seções seguintes são apresentados trabalhos com o objetivo de identificar características na fala na seção 3.1, e trabalhos com objetivo de avaliar a fluência em leitura, na seção 3.2.

3.1 Identificação de características na leitura

A avaliação de prosódia por meio de análises espectrográficas pode ter diversas abordagens, esta seção traz trabalhos que falam sobre algumas delas.

Apesar de (ARANTES, 2011) não ser um trabalho focado na avaliação de leitura de crianças, nele são apresentados alguns algoritmos interessantes para a descrição de três características prosódicas: duração, frequência fundamental *F0*, e ênfase espectral. Os algoritmos foram criados com auxílio de dados retirados do software *Praat* [1].

O algoritmo de duração descrito em (ARANTES, 2011) consiste em normalizar o tempo do áudio bruto por um *z-score* estendido, suavizar o contorno de duração normalizado por meio de uma função de média móvel de 5 pontos e por fim detectar picos de contorno de duração suavizada.

Para a frequência fundamental, são apresentados dois algoritmos (ARANTES, 2011):

- Contornos normalizados temporalmente: Consiste em captar medidas *F0* em intervalos de análise, como sílabas, vogais, palavras ou frases, e compará-las por meio de sobreposição de curvas de forma a captar padrões prosódicos com isso.

- Detecção de máximos e mínimos: Localiza pontos máximos e mínimos locais ou globais do *F0* ao longo do áudio. Pontos máximos e mínimos do *F0* indicam regiões de crescimento e decrescimento da entonação do leitor avaliado pelo algoritmo.

Por fim, o algoritmo desenvolvido para a análise da Ênfase Espectral é calculada por meio da diferença da intensidade acústica total de um sinal com a intensidade acústica do mesmo sinal, sendo este modificado por um filtro passa-baixas com limite de bandas definido pela expressão $1.5F0$.

O trabalho desenvolvido em (BAJO; FARRÚS; WANNER, 2016) também utiliza o *Praat* como software auxiliar para a análise de prosódia. Nele é desenvolvido um esquema que auxilia na representação acústica da prosódia, que consiste em um guia que indica as situações em que deve-se marcar as frases prosódicas e as palavras proeminentes dentro das frases prosódicas. As frases prosódicas são definidas por Bajo, Farrús e Wanner (2016) como unidades prosódicas que formam uma unidade homogênea em termos de F0 e curvas de intensidade. Bajo, Farrús e Wanner (2016) fizeram também um serviço web chamado de "Prosody Tagger", que recebe um texto no formato *TextGrid* e um áudio *wav* de entrada, então retorna o texto marcado com *tags* nas frases prosódicas e

[1] ⟨https://www.fon.hum.uva.nl/praat/⟩

nas palavras proeminentes de cada frase, com *features* acústicas do áudio. Este trabalho também não tem como foco a avaliação de leitura de crianças, mas é interessante analisar os parâmetros utilizados por eles para classificar padrões de prosódia, já que eles ajudam no desenvolvimento de modelos de treinamento focados na análise de prosódia.

Em (TEIXEIRA; BARBOSA; RASO,), *Praat* é utilizado também como ferramenta para extração de parâmetros acústicos em cada fronteira prosódica, desde que tenha sido indicada por no mínimo 7 dos 14 anotadores chamados para segmentar 11 textos de fala espontânea em trechos separados por fronteiras prosódicas identificadas e marcadas pelos anotadores. Foram utilizados dois métodos de classificação estatística para "gerar modelos com subconjuntos de parâmetros acústicos, que poderiam funcionar como preditores de fronteiras prosódicas" (TEIXEIRA; BARBOSA; RASO,). Para cada segmento prosódico, são extraídas 11 medidas acústicas globais e locais, essas medidas são então passadas para dois modelos estatísticos de classificação, o *Random Forest* (RF) e o *Linear Discriminant Analysis* (LDA). São utilizados para "identificar a combinação de medidas que melhor explicam a segmentação realizada pelos segmentadores perceptualmente". Os dois modelos consideram ausência ou presença de fronteira para fronteiras terminais e não terminais.

Já (QIAN et al., 2010; SRIDHAR; BANGALORE; NARAYANAN, 2008) utilizam um padrão de marcação de eventos prosódicos chamado ToBI para detecção de quebras e acentos tônicos. As quebras indicam índices de disjunção entre cada par de palavras, índices indo de 0 a 4, sendo 0 a ausência de separação, ou clitização, e 4 uma pausa completa. (QIAN et al., 2010) a utilizam em um contexto de reconhecimento de fala independente e dependente do falante. Utilizam o texto gerado para predizer rótulos prosódicos e para detectar eventos prosódicos considerando também o sinal acústico. Há também casos como em (SRIDHAR; BANGALORE; NARAYANAN, 2008) que utilizam ToBI para criação de um *framework* de marcação automática de prosódia baseada em entropia máxima. A abordagem de máxima entropia oferecida por (SRIDHAR; BANGALORE; NARAYANAN, 2008) é capaz de modelar as incertezas em rótulos, o que segundo o trabalho, auxilia na detecção de prosódia dada a ambiguidade presente na representação de eventos prosódicos através de rótulos categóricos.

3.2 Avaliação de leitura

Em (BOLAÑOS et al., 2013a; BOLAÑOS et al., 2013b), a avaliação do NAEP foi feita por meio de *features* que serviram como referência para um sistema de aprendizado de máquina. É importante ressaltar que a avaliação do NAEP utilizada nesses trabalhos é um pouco diferente da apresentada em (BENJAMIN et al., 2013), possuindo apenas 4 níveis de prosódia, o qual o autor relaciona com fluência: os níveis 1 e 2 indicam leitores não fluentes, enquanto os níveis 3 e 4 indicam leitores fluentes. As *features* prosódicas têm grande influência nos níveis do NAEP, pois são indicadores da fluência em leitura (BOLAÑOS et al., 2013a). Por exemplo, de acordo com Bolaños et al. (2013a), as *features* prosódicas P5, P6, P7 e P8, as quais são relacionadas ao número e duração das pausas, estão ligadas ao comprimento dos agrupamentos de palavras, que desempenha um papel importante na definição dos níveis do NAEP. Os autores relacionam as *features* ao que estão ligadas e como isso impacta na fluência, para cada uma delas.

Em (BOLAÑOS et al., 2013a) é explorada uma abordagem de classificação de leitura por meio de um classificador *Support Vector Machine* (SVM). A avaliação da fluência em si se dá por meio das características do áudio. Em (BOLAÑOS et al., 2013a), foram determinadas diversas *features* representando diferentes propriedades a serem analisadas em leituras. Cada *feature* recebeu um peso w_i utilizado na função de decisão do SVM, $D(x) = w.x + b$. As *features* utilizadas são divididas entre léxicas, mais relacionadas a medidas de acurácia e velocidade de leitura, e prosódicas, mais relacionadas a medidas de expressividade. As *features* léxicas e prosódicas, sendo as prosódicas as mais interessantes ao presente trabalho, são as seguintes:

- *Features* Léxicas:
  - Palavras corretas por minuto (L1): Número de palavras lidas corretamente em um minuto. Expressa a taxa de leitura. A palavra é considerada errada quando o reconhecedor não a entende ou quando são puladas;
  - Palavras lidas por minuto (L2): Número de palavras lidas em um minuto. Se difere da L1 por contar o número de palavras reconhecidas no total;
  - Número de repetições (L3): Conta as palavras repetidas cuja repetição não

está presente no texto lido pelo estudante. A maioria das repetições ocorreram na escala de uma ou duas repetições. Uma repetição diminui o NAEP de um leitor dado que na maioria das vezes ela ocorre pois o estudante leu errado a palavra e está tentando corrigi-la ao repetir a leitura;

- Número de *trigram-backoffs* (L4): Ocorre quando um leitor reconhece uma palavra incorretamente ou insere uma palavra nova no texto;
- Variância da taxa de leitura de sentenças (L5): Um bom leitor tem aproximadamente a mesma velocidade de leitura para cada palavra em um texto. Quando o leitor não é bom, o tempo lendo cada palavra pode variar devido à sua dificuldade, o que produz maior variância na taxa de leitura da sentença. Existem trabalhos que associam esse fenômeno à dificuldade de reconhecimento de algumas palavras.

- *Features* Prosódicas:
  - P1: Diferença entre o número de vezes que percebe-se uma pausa na leitura em regiões do texto marcadas como "fim de frase" e o número de pontuações presentes no texto que indicam "fim de frase";
  - P2: Diferença entre o número de vezes que percebe-se uma pausa na leitura em regiões do texto marcadas como "pontuações no meio da frase" e o número de pontuações presentes no texto que indicam "pontuações no meio da frase";
  - P3: Tamanho médio de regiões silenciosas com o número de pontuações presentes no texto;
  - P4: Diferença entre o tamanho de regiões silenciosas motivadas por pontuação e entre regiões silenciosas no áudio em geral. Essa métrica visa diferenciar pausas de decodificação e pausas de pontuação;
  - P5: Número médio de palavras entre regiões silenciosas;
  - P6: Número de regiões silenciosas;
  - P7: Duração média de regiões silenciosas;

  - P8: Duração da maior região silenciosa. Regiões silenciosas longas podem indicar tentativa de decodificação de palavras;
  - P9: Número de pausas preenchidas;
  - P10: Duração média das pausas preenchidas;
  - P11: Duração da maior pausa preenchida;
  - P12: Tamanho médio da duração de pronúncia de uma sílaba;
  - P13: Tamanho máximo da duração de pronúncia de uma sílaba;
  - P14: Diferença da média de tons de sílabas destacadas das sílabas não destacadas;
  - P15: Diferença da média de duração de sílabas destacadas das sílabas não destacadas.

Essas *features* podem ser úteis para a avaliação da prosódia como um todo em um momento posterior, dado que é necessário que se tenha informações sobre as pausas e palavras identificadas automaticamente no sinal de fala. O presente trabalho então contribui para uma avaliação parcial da prosódia, pois algumas *features* são geradas a partir da informação das pausas.

Todos os próximos trabalhos abordados nesta seção têm como foco a avaliação de leitura de crianças. Por exemplo, em (DUONG; MOSTOW; SITARAM, 2011), a avaliação de leitura é feita em cima de quatro aspectos: Expressividade, Fraseamento, Suavização e Ritmo. As métricas foram calculadas a partir de duas *features* prosódicas e três *features* léxicas:

- *Features* Prosódicas:
  - CRi: Calculada a partir dos contornos rítmicos da análise espectrográficas;
  - CFi: Calculada a partir dos contornos de Tom, *F0*, da análise espectrográficas;
- *Features* Léxicas:
  - CWPM: Palavras lidas corretamente por minuto;
  - IWPM: Palavras omitidas, lidas incorretamente e repetidas por minuto;

– VPM: Vogais lidas por minuto;

Esse trabalho se aproxima do de Bolaños et al. (2013a) na medida em que também apresenta características léxicas e prosódicas. As *features* prosódicas em questão são obtidas a partir da projeção do áudio da leitura de crianças na leitura de adultos. As *features* das leituras de adultos foram obtidas a partir de escalas multidimensionais (MDS) que detectam padrões prosódicos dos áudios que analisa. A expressividade da leitura das crianças é medida de duas formas: a primeira a partir da similaridade da leitura de um texto por um estudante com a leitura do mesmo texto por um adulto. A segunda trata-se de um modelo generalizado, treinado a partir de leituras de adultos, que ajuda na avaliação da leitura das crianças. O modelo foi baseado em modelos de duração, frequência fundamental (F0) e intensidade para mapear texto para prosódia, porém, foram usados para avaliar a prosódia infantil ao invés de serem usados para prescrever um contorno prosódico específico.

Em (SABU; RAO, 2021), a avaliação de prosódia se dá por medidas estatísticas sobre quatro famílias de *features*: frequência fundamental, intensidade, duração e ênfase espectral. Essas *features* então são submetidas a florestas classificadoras aleatórias que analisam a presença de quebras de frases e palavras proeminentes nos áudios.

3.3 Considerações Finais

A Tabela 3.1 traz a relação entre os trabalhos apresentados e pontos importantes para o presente trabalho, de forma a auxiliar na análise quanto às diferenças e semelhanças de cada trabalho, inclusive com o presente trabalho, a fim de comparação e análise dos diferenciais de cada. Os trabalhos citados neste capítulo apresentaram formas de se identificar características da fala, assim como avaliar fluência e prosódia na leitura, seja em leituras feitas por crianças ou não. Mas não foram encontrados trabalhos cujo objetivo de análise tenha sido a pausa na leitura, apenas trabalhos que abordem ela como parte da avaliação da prosódia na leitura em voz alta. Nesse sentido, o presente trabalho contribui para a literatura, pois pretende trazer um enfoque em um componente específico e importante da prosódia, a pausa de sentido.

Tabela 3.1: Comparação metodológica dos trabalhos relacionados

Trabalho	***Features*** **léxicas**	***Features*** **espectográficas**	***Features*** **prosódicas**	**Leitura de crianças**	**Avaliação da prosódia**	**Avaliação das pausas de sentido**
Presente Trabalho	-	X	X	X	-	X
(ARANTES, 2011)	-	-	X	-	-	-
(BAJO; FARRÚS; WANNER, 2016)	-	-	X	-	X	-
(TEIXEIRA, 2018)	-	-	X	-	-	-
(SRIDHAR; BANGALORE; NARAYANAN, 2008)	-	-	X	-	X	-
(BOLAÑOS et al., 2013a)	X	-	X	-	X	-
(DUONG; MOSTOW; SITARAM, 2011)	X	-	X	X	X	-
(SABU; RAO, 2021)	-	-	X	X	X	-

4 Materiais e Métodos

O presente capítulo trata da composição da base de dados usada para a realização dos experimentos e os métodos utilizados para avaliação das informações dos áudios disponíveis. Os detalhes relacionados ao conjunto de dados utilizados são abordados na seção 4.1, trazendo desde a quantidade de áudios até as informações utilizadas de cada áudio. A seção 4.2 apresenta detalhes quanto ao modelo de reconhecimento automático de fala (ASR) utilizado para o reconhecimento automático das leituras de cada áudio. Por fim, a seção 4.3 apresenta o *workflow* com todas as etapas da realização dos experimentos.

4.1 Base de dados

A base de dados utilizada no trabalho é composta por 1039 áudios de leitura de texto em voz alta realizadas por crianças em fase de alfabetização. A distribuição de fluência e de respeito às pausas de acordo com a avaliação manual dessa base pode ser observada na Tabela 4.1 e 4.2 respectivamente. Esse conjunto de áudios é uma amostra representativa selecionada aleatoriamente de uma base de 55 mil áudios obtidos através de avaliações de leitura realizadas pela Fundação CAEd/UFJF, que consistem de leituras de voz alta de crianças do Ensino Fundamental I, as quais foram instruídas a ler um texto narrativo composto de 250 palavras. Durante as avaliações de leitura, foram gravados apenas o primeiro minuto de leitura de cada áudio, pois é o suficiente para uma avaliação da primeira impressão da leitura.

A amostra foi definida para representar a base original com 95% de grau de confiança e 3% de margem de erro. Todos os áudios dessa base foram avaliados manualmente por corretores especialistas em alfabetização. Para cada áudio, os avaliadores informaram a última palavra do texto lida pelo falante, quantidade de palavras lidas corretamente, se a leitura como um todo respeitou ou não as pausas, entre outras informações. Das informações disponíveis para cada áudio, não foram passadas quaisquer informações a respeito da criança que está lendo o texto.

Para o presente trabalho, apenas a informação quanto ao respeito ou não das pausas durante a leitura é importante, pois ela é utilizada para fins de comparação dos resultados automáticos. A quantidade total de áudios foi separada em duas bases: base de investigação e base de teste. A base de investigação é composta por 79 áudios, divididos entre leitores que respeitam a pausa (52 áudios) e que não respeitam (27 áudios), sendo assim uma base desbalanceada, conforme mostrada na Tabela 4.3.

A escolha de uma base de investigação foi feita para gerar estatísticas preliminares sobre dados acústicos dos áudios, sendo que a seleção desses 79 áudios se deu pela escolha de áudios cuja leitura foi totalmente certa de acordo com o reconhecimento automático da qualidade da leitura de cada palavra. Foi importante que a base de investigação fosse composta apenas de áudios com leituras corretas para evitar problemas nas primeiras análises das pausas por conta de alinhamentos equivocados. Já a base de teste é composta por 859 áudios, que restaram da amostra de 1039 após retirada dos áudios sem pausa alguma, distribuídos conforme observado na Tabela 4.4. Essa base foi utilizada para testar o modelo com as melhores *features* observadas na base de investigação. Para a avaliação das pausas de forma adequada apesar dos erros de leitura presentes na base de teste, foi decidido que o alinhamento seria feito de forma a ignorar as leituras erradas, considerando apenas pausas entre palavras lidas corretamente.

Tabela 4.1: Distribuição de fluência da base de áudios

Fluência	Quantidade de áudios
Fluentes	521
Disfluentes	518
Total	1039

Tabela 4.2: Distribuição de respeito às pausas da base de áudios

Respeito às pausas	Quantidade de áudios
Respeito	343
Desrespeito	696
Total	1039

Tabela 4.3: Distribuição de áudios da base de investigação

Respeito às pausas	Quantidade de áudios
Respeito	52
Desrespeito	27
Total	79

Tabela 4.4: Distribuição de áudios da base de teste

Respeito às pausas	Quantidade de áudios
Respeito	289
Desrespeito	570
Total	859

4.2 Modelo ASR

Para a classificação automática em relação ao respeito às pausas na leitura, é utilizado um alinhamento com base no resultado de um sistema de reconhecimento automático de fala (ASR). ASR consiste na utilização de inteligência artificial para gerar texto transcrito a partir dos sinais sonoros de uma fala.

O modelo utilizado neste trabalho para o reconhecimento das falas dos áudios consiste em um ajuste fino (*fine-tuning*) do modelo *Wav2Vec2*, um modelo pré-treinado para reconhecimento de fala, cujo *framework* pode ser observado na Figura 4.1.

Normalmente, os modelos ASR tendem a tentar identificar corretamente um sinal de fala dentro do léxico de uma língua, passando assim para texto. Mas o *Wav2Vec2* tem o diferencial de considerar o contexto em que o sinal de fala está inserido, pois ele é treinado "mascarando" (ocultando) um segundo de uma parte do áudio (5 segundos) e tentando descobrir o que foi dito nesse segundo com base no contexto em volta. Assim, o modelo gera uma representação vetorial de cada sinal de fala, de forma a relacionar as falas entre si e descobrir o que foi dito naquele segundo com base na transcrição do áudio naquele momento.

O *fine-tuning* utilizado no presente trabalho treinou o modelo para a tarefa de reconhecimento de fala em língua portuguesa. Por questões de *hardware* limitado, foi utilizado um modelo disponibilizado pela comunidade científica (JUNIOR et al., 2021), o qual foi treinado por 20 épocas com aproximadamente 290 horas de áudios na língua portuguesa acompanhados de suas transcrições. Apesar dos áudios serem de falantes adultos, a base de treinamento se mostrou suficientemente grande para generalizar o reconhecimento, sendo adequado também para o reconhecimento de falas de crianças.

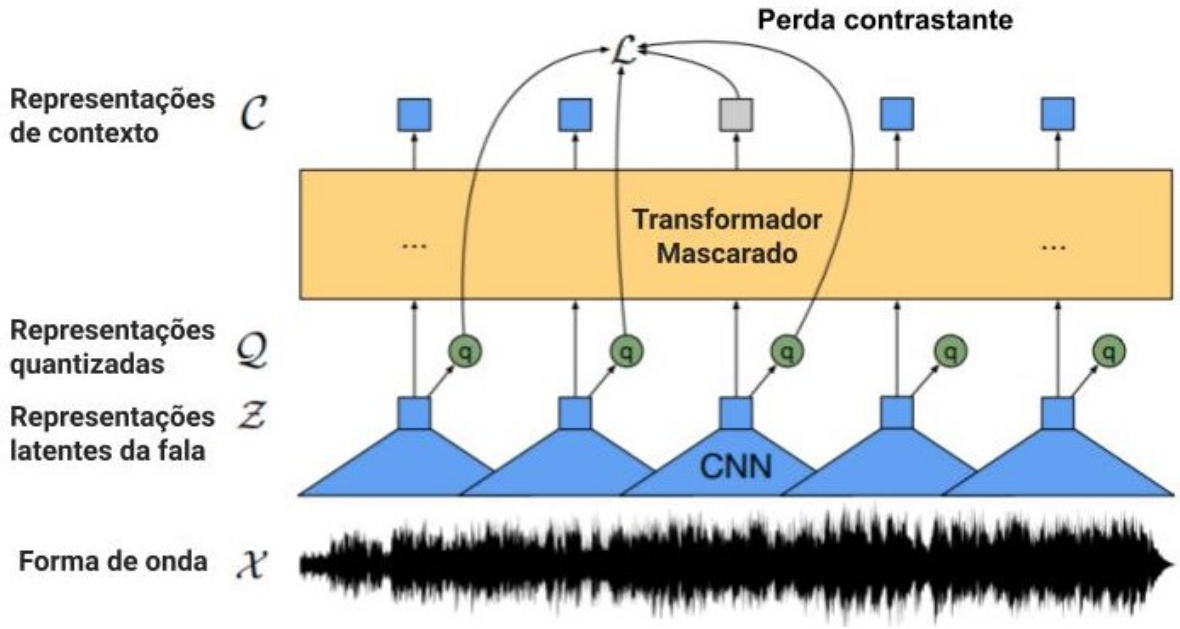

Figura 4.1: Ilustração da estrutura do *wav2vec2* que aprende tanto representações de fala contextualizadas quanto um inventário de unidades de fala discretizadas.

Como o reconhecimento da fala é feito na ordem do texto esperado de ser lido, é possível relacionar as probabilidades de leitura de cada letra do texto de referência com a qualidade da leitura definindo uma pontuação a partir da probabilidade. Após obter o texto transcrito com base no áudio de leitura textual e as pontuações de cada letra, é realizado um alinhamento com o texto de referência, de forma a armazenar informação da qualidade de leitura de cada palavra e de cada letra, conseguindo definir automaticamente se a palavra foi lida corretamente ou não, a partir de um limiar. Essas informações então são passadas como entrada para o algoritmo que define as métricas usadas nos experimentos apresentados no Capítulo 5.

4.3 Fluxo de realização dos experimentos

O processo completo dos experimentos podem ser visualizados na Figura 4.2, em que é apresentado um fluxograma de todo o processo, dividido em etapas. A primeira etapa consiste em selecionar os áudios que participarão dos experimentos. Então eles são passados juntamente com seus respectivos textos de referência para o sistema de reconhecimento automático de fala, de forma que o alinhador gera as probabilidades para cada letra do texto de ter sido lida e os milissegundos de início e fim da leitura. As pausas identificadas a partir do momento de início e fim da leitura de cada palavra então são tratadas de forma a terem seus *outliers* removidos.

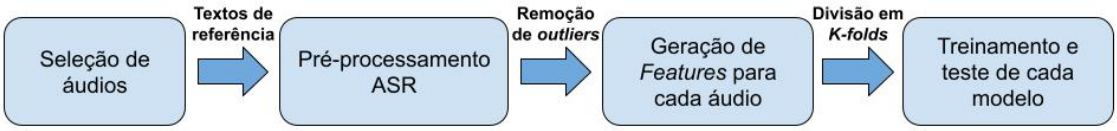

Figura 4.2: *Workflow* dos experimentos realizados

Para cada áudio são geradas as *features* escolhidas para os testes e armazenado seu resultado esperado de acordo com a classificação manual, de forma a verificar o resultado. A base com as métricas representado cada áudio então é dividida em 10 blocos, cada um respeitando a distribuição de classificação da base de teste completa. Por fim, para cada um dos blocos, os modelos são treinados com 2/3 da quantidade de áudios e testados com 1/3 restante, de forma que a média das acurácias obtidas indica de forma mais confiável o desempenho de cada modelo, como forma de tratar qualquer impacto que a aleatoriedade em que os áudios são separados possa causar.

5 Experimentos

Com o objetivo de analisar as pausas dos áudios da base de dados e verificar o estado dos áudios, sua distribuição e métricas possíveis de serem utilizadas, foi feita uma análise dos dados extraídos da base de investigação. Em um primeiro momento, o arquivo de saída do reconhecimento automático de fala e alinhamento dos áudios foi analisado a fim de entender quais informações poderiam ser utilizadas. Dos dados obtidos, foram utilizados as informações sobre acerto na leitura de cada palavra, a classificação da leitura quanto ao respeito às pausas e os tempos de início e fim da leitura de cada palavra, identificando assim as pausas entre palavras lidas corretamente. Ao observar as classificações manuais em relação às pausas nos áudios, também foi possível perceber que o critério de classificação do avaliador quanto ao respeito às pausas do texto como um todo é subjetiva, sem qualquer critério quantitativo para classificar a leitura, o que implica em uma dificuldade maior na classificação automática com alta taxa de acerto.

Apenas dados relacionados à descoberta das pausas se fez necessário, dado que o objetivo final do trabalho é acertar o máximo possível a classificação da leitura da criança, entre as que respeitaram as pausas e as que não as respeitaram. Mas para verificar a relevância do trabalho, primeiro foi analisado se o QPL (quantidade de palavras lidas) que o corretor indicava para cada áudio, tinha alguma relação com sua classificação de respeito às pausas. Não foram encontradas dependências entre as informações, conforme observado na Figura 5.1, o que mostrou que não há forma de concluir a classificação das pausas utilizando alguma outra métrica fornecida.

Após descobrir a distribuição dos áudios de acordo com sua classificação e com os resultados do ASR sobre qualidade da leitura, conforme mostradas nas Tabelas 4.1 e 4.2, foram analisadas as durações de cada tipo de pausa dos áudios. Ao analisá-las, foi observada a existência de *outliers* e que sua presença poderia atrapalhar a conclusão sobre as melhores métricas, fazendo-se necessária a sua retirada. Esses *outliers* nada mais são do que pausas normais durante a leitura que possuem duração muito diferente da maioria, o que pode prejudicar a média das pausas. Os *outliers* representam momentos em que

a criança demora mais para continuar a leitura, seja por uma respiração mais cumprida, por nervosismo por conta da avaliação ou até dúvida em relação à pronúncia. O problema é que esse tipo de pausa muito diferente das demais na maioria das ocasiões não é uma pausa de sentido e sim de outro fator, o que foge do escopo a ser avaliado pelo trabalho, o qual consiste na avaliação quanto ao respeito às pausas de sentido durante a leitura. Isso demonstra a necessidade de desconsideração das pausas *outliers*.

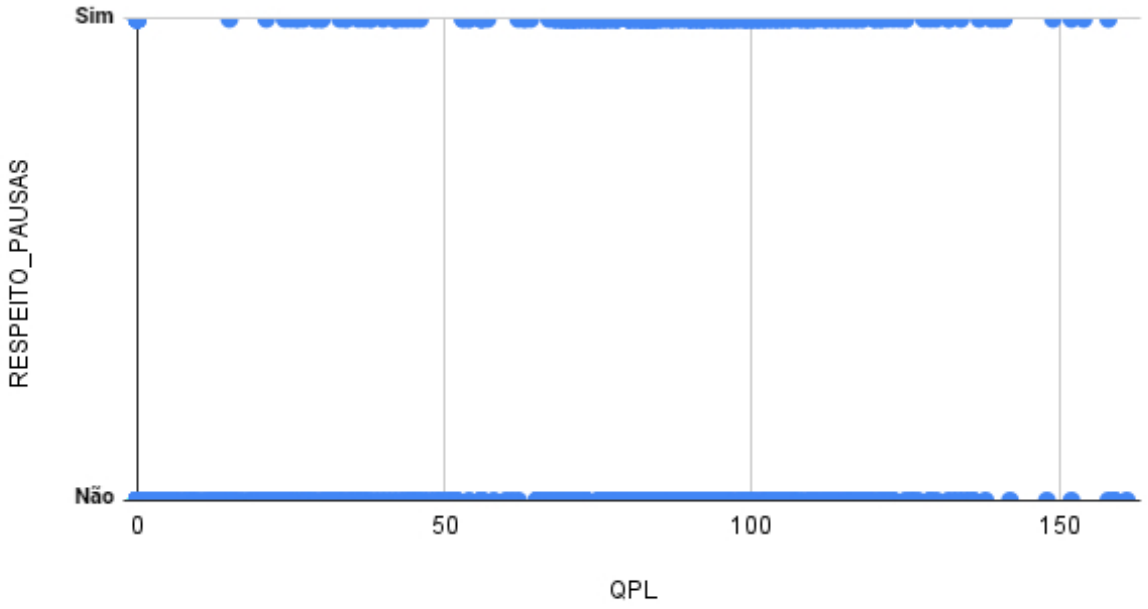

Figura 5.1: Gráfico de dispersão de valor do QPL para cada áudio de acordo com sua classificação de respeito às pausas durante a leitura.

A fim de analisar com mais detalhes a presença dos *outliers* e a relação entre os tipos de pausas em cada áudio, foram gerados gráficos de linha para cada um dos 79 áudios, para verificar a relação das durações das pausas em milissegundos ao longo do áudio. Um exemplo desse gráfico para um áudio que respeitou as pausas com a presença de *outliers* pode ser observado na Figura 5.2, enquanto o mesmo áudio sem a presença de *outliers* pode ser observado na Figura 5.3.

É possível observar na Figura 5.2 que os *outliers* presentes nas pausas normais ficam mais próximos da duração das pausas de vírgulas e de pontos, o que atrapalha na classificação com base nesses valores, dado que não são tão diferentes assim entre si. Após a retirada dos *outliers*, conforme apresentado na Figura 5.3, a discrepância entre as pausas normais e as pausas por pontuação se tornou muito mais perceptível, permitindo assim,

a possibilidade da definição de algum tipo de limiar ou critério de classificação utilizando a relação entre os diferentes tipos de pausas.

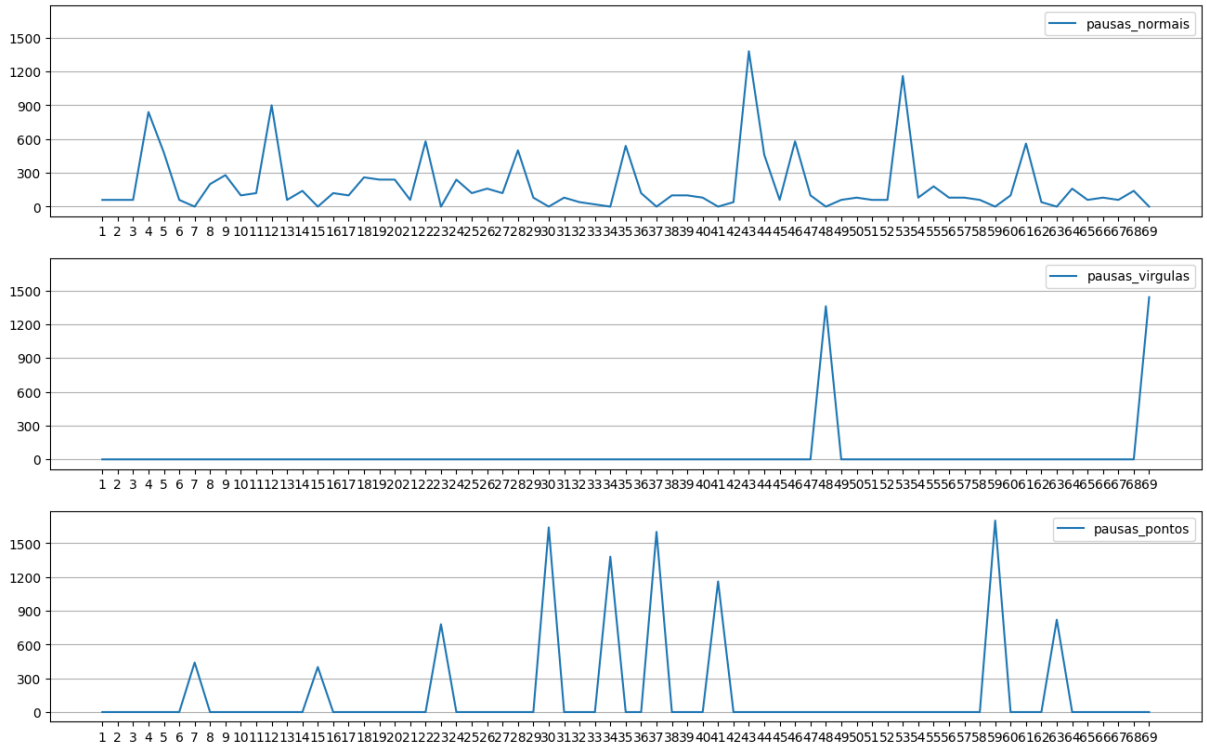

Figura 5.2: Gráfico de linha com *outliers* de Duração da pausa (ms) x Número da pausa para um áudio que respeitou as pausas de acordo com avaliador manual.

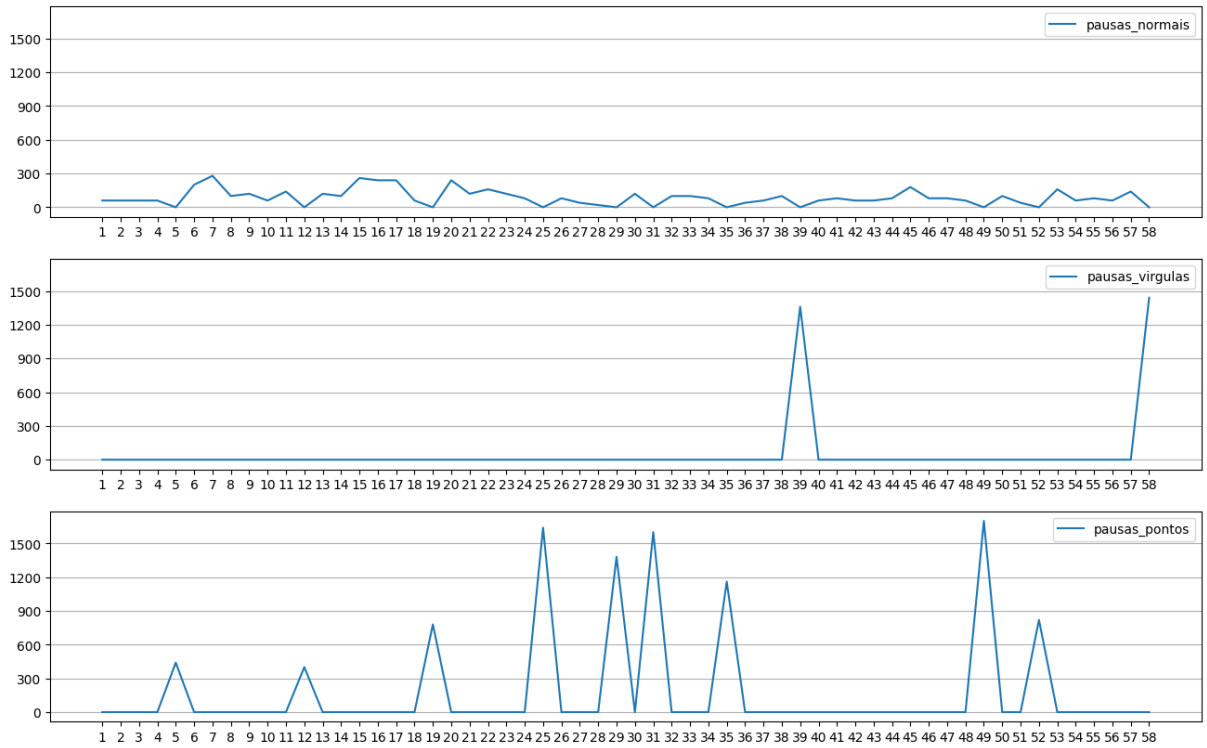

Figura 5.3: Gráfico de linha sem *outliers* de Duração da pausa (ms) x Número da pausa para um áudio que respeitou as pausas de acordo com avaliador manual.

Foram então gerados *boxplots* sobre as pausas dos áudios e sua classificação quanto ao respeito às pausas, com o objetivo de identificar diferenças entre as leituras que respeitaram e as que não respeitaram as pausas, demonstrando a possibilidade de encontrar um padrão para classificação. Um dos *boxplots* gerados, é apresentado na Figura 5.4, em que é possível observar o *boxplot* da razão das medianas das pausas de vírgulas e de pontos em relação à mediana das pausas normais para cada áudio. Foi utilizada a razão entre as medianas pois é importante verificar a relação entre as durações das pausas de vírgulas e de pontos quando comparadas com a duração do restante das pausas durante a leitura, dado que cada criança tem sua velocidade de leitura.

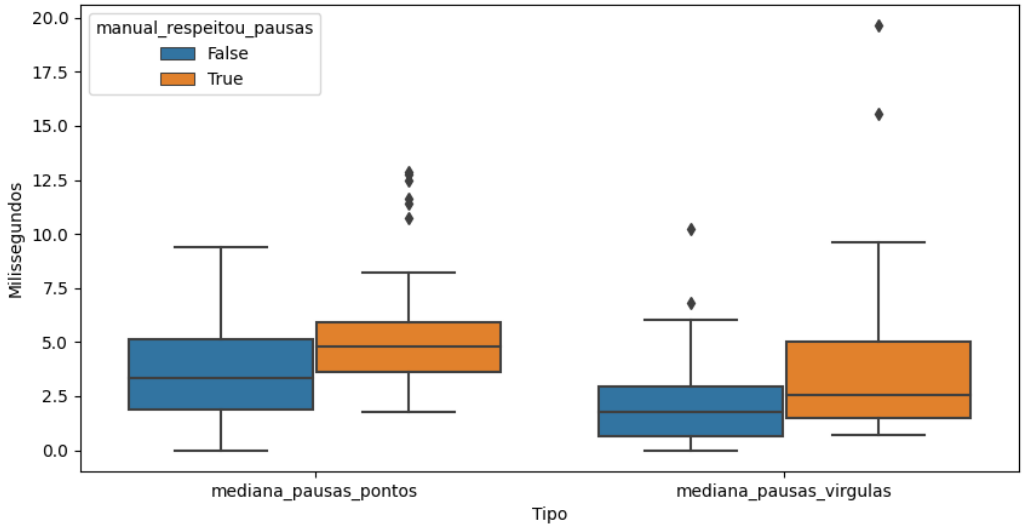

Figura 5.4: Gráfico de *boxplot* com as razões das medianas de pausa de vírgula e de ponto com a mediana das pausas normais de cada áudio.

Conforme observado na Figura 5.4, os *boxplots* laranjados se referem aos áudios cuja classificação manual foi de que respeitaram as pausas, enquanto os azuis são o contrário. Eles foram gerados a partir das razões entre cada mediana por tipo, com a mediana das pausas normais. Por exemplo, no caso das pausas de pontos, à esquerda na Figura 5.4, foi calculada a razão entre a mediana das pausas de ponto com a mediana das pausas normais de seu respectivo áudio, para cada um dos áudios presentes na base. Ao comparar os *boxplots* dos dois tipos de classificação, é possível perceber que a mediana dos áudios que respeitaram as pausas são quase do mesmo valor que o terceiro quartil dos que não respeitaram. Ou seja, a possibilidade da existência de uma regra de classificação quanto ao respeito às pausas se mostra promissora, dado que há diferença no

comportamento dos *boxplots* de classificações diferentes.

5.1 Definição das *features*

Para utilizar de aprendizado de máquina e treinar diferentes modelos de classificação dos áudios, se faz necessário que a máquina seja treinada com base em alguma ou algumas informações sobre cada áudio. Essas informações extraídas dos áudios, são as métricas ou *features* utilizadas como parâmetro para o treinamento da máquina responsável por classificar automaticamente os áudios. É importante que as métricas utilizadas consigam representar bem os áudios e permitir que a máquina consiga perceber algum padrão para classificar corretamente os áudios. Todas as *features* criadas estão apresentadas na Tabela 5.1 com suas respectivas definições.

As *features* foram criadas também pois a Fundação CAEd/UFJF, responsável por disponibilizar os áudios e suas avaliações manuais, não possuem um padrão definido ainda para a classificação quanto ao respeito às pausas de sentido, sendo uma classificação mais subjetiva da percepção do avaliador humano. A Fundação CAEd/UFJF. Então é interessante que a Fundação tenha informações mais granulares quanto às pausas dos áudios, de forma a ser possível uma definição de uma regra de classificação com base nessas informações, as quais são as *features* criadas. Essa granularidade permite também uma flexibilidade na definição das regras de classificação, podendo mudar a qualquer momento de forma mais simples. A Fundação já possui a preferência a dados mais granulares, pois as outras classificações de fluência utilizadas pela empresa já dependem de dados mais detalhados.

Tabela 5.1: Definição das *features* utilizadas para testar os modelos de classificação

Feature	**Definição**
F_{med}	Medianas das pausas de cada áudio
F_{mean}	Médias das pausas de cada áudio
F_{mm}	$F_{med} \cup F_{mean}$
F_{med_d}	Diferença das medianas das pausas de vírgula e ponto com a mediana das pausas normais
F_{mean_d}	Diferença das médias das pausas de vírgula e ponto com a média das pausas normais
F_{mm_d}	$F_{med\_d} \cup F_{mean\_d}$
F_{med_r}	Razão das medianas das pausas de vírgula e ponto com a mediana das pausas normais
F_{mean_r}	Razão das médias das pausas de vírgula e ponto com a média das pausas normais
F_{mm_r}	$F_{med\_r} \cup F_{mean\_r}$
$F_{d\_q2v\_q3n}$	Diferença da mediana das vírgulas e o terceiro quartil das pausas normais
$F_{r\_q2v\_q3n}$	Razão da mediana das vírgulas e o terceiro quartil das pausas normais
$F_{d\_q2p\_q3n}$	Diferença da mediana dos pontos e o terceiro quartil das pausas normais
$F_{r\_q2p\_q3n}$	Razão da mediana dos pontos e o terceiro quartil das pausas normais
$F_{d\_q2vp\_q3n}$	$F_{d\_q2v\_q3n} \cup F_{d\_q2p\_q3n}$
$F_{r\_q2vp\_q3n}$	$F_{r\_q2v\_q3n} \cup F_{r\_q2p\_q3n}$
F_{all_r}	$F_{mm\_r} \cup F_{r\_q2vp\_q3n}$
F_{all_d}	$F_{mm\_d} \cup F_{d\_q2vp\_q3n}$
F_{all}	$F_{all\_r} \cup F_{all\_d}$

As métricas devem ser valores que representem o áudio como um todo, pois a classificação que a máquina terá de fazer será sobre o áudio inteiro. Portanto, foram utilizadas *features* como o valor da mediana das pausas (F_{med}), suas médias (F_{mean}),

entre outras, conforme observado na Tabela 5.1.

Levando em consideração que os ritmos e velocidade de leitura variam em cada áudio, foram definidas também métricas que consideram comparações entre as pausas de pontuações e as pausas normais do mesmo áudio. Uma das métricas criadas pensando em avaliar essa relação, foi a diferença entre as medianas das pausas de vírgulas e de pontos com a das pausas normais (F_{med_d}).

Dada a quantidade de experimento e a necessidade de informações úteis e de diferentes tipos para que a máquina consiga descobrir como relacioná-las e classificar o áudio com base nelas, foram consideradas para os testes com a base de teste apenas as *features* que são a união de outras *features*, o que significa que elas contém as informações presentes nas *features* que as compõem. Como por exemplo a $F_{d\_q2vp\_q3n}$, F_{all_d}, entre outras, totalizando 8 métricas, como é possível observar na Tabela 5.2, na seção de Resultados.

5.2 Resultados

Para realizar os experimentos, foi utilizado o método de divisão da base de teste (definida na Tabela 4.4) chamado de *StratifiedKFold*, em que a base é dividida em partes chamadas de *folds* de forma a manter a distribuição dos áudios de acordo com a classificação. Para cada *fold*, é usado um método de divisão entre base de treino e de teste, de forma que o modelo de classificação é treinado com a base de treino do respectivo *fold* e então tem sua acurácia testada usando a base de teste equivalente.

Quanto aos modelos de classificação utilizados, foram realizados experimentos com o *Random Forest Classifier* (RFC), *Logistic Regression* (LR), *Voting Classifier* (VC), *XGBoost* (XGB) e *MLP Classifier* (MLP). O VC utilizou como classificadores o RFC, LR e o *Gaussian Naive Bayes*, conforme explicado no Capítulo de Fundamentação Teórica. Dos classificadores utilizados, o único com nenhuma mudança nos parâmetros foi o *Gaussian Naive Bayes*, por isso sua ausência na Tabela 5.2, a qual é responsável por mostrar os parâmetros utilizados para cada modelo. É importante que se entenda que os parâmetros foram definidos de forma arbitrária com base em alguns resultados obtidos com a base de investigação, exceto o parâmetro "solver" que foi escolhido o recomendado para bases maiores. Como a distribuição e comportamento da base de investigação difere da base de

teste, é provável que existam parâmetros mais adequados para a base de teste.

Os resultados quanto à acurácia máxima obtida para cada modelo em cada uma das *features* testadas estão evidenciados na Tabela 5.3. O resultado é em relação à acurácia máxima pois são obtidas várias acurácias, uma para cada um dos *folds* usados para treinar e testar os modelos. Então a acurácia máxima significa o *fold* em que o modelo foi melhor treinado e obteve os melhores resultados.

Tabela 5.2: Parâmetros utilizados nos experimentos para cada um dos modelos

Modelos	**Parâmetros**		
	n_estimetors	**max_features**	**max_depth**
Random Forest Classifier	1200	"sqrt"	540
	multi_class	**solver**	**max_iter**
Logistic Regression	"ovr"	"saga"	500
	voting		**n_jobs**
Voting Classifier	"soft"		8
	objective		
XGBoost	"binary:logistic"		
	activation	**solver**	**max_iter**
MLP Classifier	"relu"	"adam"	800

Tabela 5.3: Resultados de acurácia média com desvio padrão dos experimentos utilizando 8 *features* e 5 modelos de classificação com a base de teste

Feature	**RandomForest Classifier**	**Logistic Regressor**	**Voting Classifier**	**XGBoost**	**MLP Classifier**
F_{mm}	$67,7\% \pm 3,2\%$	$67,3\% \pm 2,5\%$	$67,8\% \pm 5,2\%$	**68,5%** $\pm 3,8\%$	$66,6\% \pm 3,3\%$
F_{mm_d}	$67,5\% \pm 4,4\%$	$63,6\% \pm 3,7\%$	$67,6\% \pm 3,7\%$	**68,2%** $\pm 4,0\%$	$65,0\% \pm 3,5\%$
F_{mm_r}	$67,2\% \pm 3,4\%$	$67,5\% \pm 3,6\%$	$67,8\% \pm 3,1\%$	$67,1\% \pm 3,4\%$	**69,0%** $\pm 5,0\%$
$F_{d\_q2vp\_q3n}$	$64,1\% \pm 2,6\%$	$64,1\% \pm 3,8\%$	$66,0\% \pm 4,2\%$	$66,7\% \pm 3,6\%$	**69,5%** $\pm 3,4\%$
$F_{r\_q2vp\_q3n}$	$66,9\% \pm 2,8\%$	$66,5\% \pm 4,0\%$	$68,0\% \pm 2,6\%$	**69,4%** $\pm 3,6\%$	$68,5\% \pm 3,7\%$
F_{all_d}	$66,6\% \pm 3,2\%$	$65,8\% \pm 2,9\%$	**68,5%** $\pm 4,3\%$	$65,9\% \pm 4,7\%$	$67,1\% \pm 4,6\%$
F_{all_r}	$66,8\% \pm 2,7\%$	$66,7\% \pm 3,0\%$	$67,8\% \pm 2,1\%$	$67,8\% \pm 3,5\%$	**69,2%** $\pm 3,9\%$
F_{all}	**68,6%** $\pm 2,3\%$	$65,9\% \pm 2,9\%$	$67,3\% \pm 4,2\%$	$67,2\% \pm 3,1\%$	$67,4\% \pm 3,5\%$

Para os experimentos com a base de teste, diferentemente da base de investigação, ocorreram leituras erradas nos áudios, então para que o alinhamento permanecesse correto na análise das pausas, foram consideradas apenas as pausas que ocorreram entre duas palavras lidas corretamente. Isso fez com que os resultados dos experimentos mostrados na Tabela 5.3 não obtivessem resultados tão bons quando comparados à base de investigação, em que no melhor dos modelos, foi obtida uma acurácia média de 81% mesmo com uma quantidade baixa de áudios. Enquanto na base de teste, o melhor modelo se mostrou ser o *MLPClassifier* em que obteve uma acurácia média de $69,5\% \pm 3,4\%$ para a *feature* $F_{d\_q2vp\_q3n}$. Vale destacar que o modelo que obteve a melhor acurácia por *fold* foi o *MLPClassifier* com a *feature* F_{mm_r}, cuja acurácia máxima alcançou 77,9%.

Após investigação da execução do algoritmo que seleciona as pausas utilizadas para definir as *features* de cada áudio, foi observado que houveram áudios em que não ocorreram pausa de vírgula e/ou de ponto por considerar apenas as pausas entre leituras corretas. Como são crianças em fase de alfabetização, a leitura incorreta é frequente, o que fez com que muitas pausas fossem desconsideradas. Essas situações de falta de algum tipo de pausa ou de ambos somaram-se em 356 áudios, equivalendo a aproximadamente 42% do total da base de teste. Com a quantidade de áudios em que isso ocorreu, o treinamento da máquina pode ter ficado prejudicado e inadequado, dado que depende das métricas passadas como parâmetros, as quais em sua maioria se tornam inválidas durante situações em que não há todos os tipos de pausas. Com o treinamento prejudicado, as predições de classificação também possuem mais chance de estarem equivocadas.

Na Figura 5.5 é apresentada a matriz de confusão do melhor resultado obtido durante os testes com as 10 partes da base de teste (*folds*). A matriz diz respeito ao resultado obtido pelo modelo *MLPClassifier* utilizando a *feature* F_{mm_r}, que apesar de não ter obtido a melhor acurácia média, obteve a melhor acurácia máxima, de 77,9%. É possível observar que o modelo tem uma dificuldade maior em classificar corretamente os casos em que a leitura respeitou as pausas de sentido, cujas precisão e revocação foram de, respectivamente, 71% e 59%. A precisão diz respeito à quantas predições o modelo acertou dentre todas as classificações de classe Positiva que fez, que no caso é a classe das leituras que respeitaram as pausas. Enquanto a revocação diz respeito a quantas classificações

o modelo acertou dentre todas as situações em que o valor esperado pertence à classe Positiva. Apesar da acurácia não ter atingido o resultado esperado, o mais preocupante é a revocação no momento, dado que é a situação com maior taxa de erro e que diz respeito às leituras esperadas de serem classificadas como as que respeitaram as pausas.

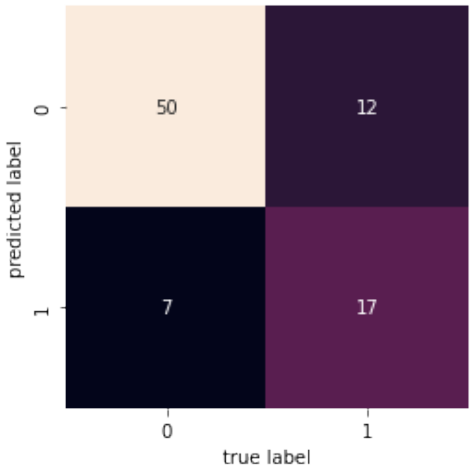

Figura 5.5: Matriz de confusão para o melhor caso - modelo *MLPClassifier* e *feature* F_{mm_r}

Independentemente dos resultados, a revocação se mostra mais importante para o problema que a precisão, na medida em que compara os acertos com apenas valores esperados de classe Positiva, enquanto a precisão considera os acertos dentro da própria classificação como pertencente à classe Positiva. Isso mostra ainda mais a importância dos próximos passos estarem voltados para a melhora na classificação do conjunto de áudios cujas leituras respeitaram as pausas.

Como falado anteriormente, as pausas de sentido compõem a prosódia, que por sua vez, faz parte do conceito de fluência em leitura, de forma que ao avaliar corretamente as pausas de sentido, indiretamente a avaliação da fluência se torna mais completa. Os resultados obtidos durante o teste, apesar de não alcançarem o esperado em um primeiro momento, se mostrou promissor para a avaliação nos casos em que o áudio foi avaliado como desrespeito às pausas, dado que o modelo acertou 88% desses casos. Com o objetivo de automatizar o processo, a avaliação deve ser o mais confiável possível. Como próximos passos então, o ideal é melhorar a classificação automática para os casos de respeito às pausas e deixar o modelo o mais confiável possível.

A utilização do classificador na prática não irá gerar gasto a mais de proces-

samento, pois as pausas são retiradas do resultado de um alinhamento automático já realizado previamente. Mas para que o classificador consiga cumprir o objetivo de poupar tempo na avaliação das leituras, se faz importante tentar descobrir características em comum entre os áudios cuja classificação foi errada. Assim, sendo possível dizer quais classificações são confiáveis e quais não, para evitar avaliar manualmente uma avaliação correta já realizada de forma automática.

Uma outra possibilidade de melhoria dos resultados pensada foi a filtragem dos áudios utilizados na base teste, retirando os áudios que não possuam um mínimo de pausas de vírgula e de pontos. Isso foi feito a fim de verificar se há alguma melhora no resultado ao ter uma quantidade maior de pausas de vírgula e ponto. Então para testar a solução, foram filtrados os áudios de acordo com uma combinação de quantidades mínimas de pausas por tipo que devem estar presentes em cada áudio. As combinações variaram entre ter pelo menos 1 pausa de vírgula presente a ter pelo menos 5 pausas de vírgula presentes e de forma análoga foi feita para as pausas de ponto. Foram gerados então 25 combinações diferentes de mínimos de pausas nos áudios, conforme apresentada na Tabela 5.4, nomeadas de forma que o número antes da letra "v" significa o mínimo de pausas de vírgula e o número antes da letra "p" o de pausas de ponto. Por exemplo, $1v\_1p$ significa que os áudios continham no mínimo uma pausa de vírgula e uma de ponto presentes, já $2v\_3p$ significa que serão desconsiderados os áudios com uma quantidade de pausas de vírgula menor que 2 e de pausas de ponto menor que 3. As combinações foram usados para testar novamente as 8 *features* criadas para cada um dos modelos utilizados, de forma a ter um total de 1000 experimentos realizados.

Tabela 5.4: Distribuição de quantidade de áudios por combinação de filtros

Combinação	**Áudios que Respeitam pausas**	**Áudios que Desrespeitam pausas**	**Total**
$1v\_1p$	254	249	503
$1v\_2p$	249	238	487
$1v\_3p$	247	226	473
$1v\_4p$	245	221	466
$1v\_5p$	244	216	460
$2v\_1p$	224	202	426
$2v\_2p$	224	201	425
$2v\_3p$	223	200	423
$2v\_4p$	221	196	417
$2v\_5p$	220	191	411
$3v\_1p$	195	163	358
$3v\_2p$	195	163	358
$3v\_3p$	195	163	358
$3v\_4p$	193	163	356
$3v\_5p$	192	160	352
$4v\_1p$	121	98	219
$4v\_2p$	121	98	219
$4v\_3p$	121	98	219
$4v\_4p$	121	98	219
$4v\_5p$	121	97	218
$5v\_1p$	50	44	94
$5v\_2p$	50	44	94
$5v\_3p$	50	44	94
$5v\_4p$	50	44	94
$5v\_5p$	50	44	94

Tabela 5.5: Melhores resultados de acurácia nos experimentos

Modelo	**Combinação**	***feature***	**Acurácia Média**
Logistic Regression	4v_5p	F_{mm_d}	60.1%
Logistic Regression	4v_5p	$F_{d\_q2vp\_q3n}$	59,2%
Logistic Regression	4v_3p	$F_{d\_q2vp\_q3n}$	58,5%
Logistic Regression	4v_4p	$F_{d\_q2vp\_q3n}$	58,5%
Logistic Regression	4v_1p	$F_{d\_q2vp\_q3n}$	58,5%
Logistic Regression	4v_2p	$F_{d\_q2vp\_q3n}$	58,5%
Voting Classifier	3v_4p	F_{all_d}	58.4%
XGBoost	4v_5p	$F_{r\_q2vp\_q3n}$	58.3%

Os experimentos mostraram que o modelo que obteve os melhores resultados foi o *Logistic Regression*, com as 6 melhores acurácias encontradas, todas acima de 58% de acurácia média, conforme observado na Tabela 5.5. O restante das informações quanto às acurácias médias obtidas em cada um dos 200 experimentos para os 5 modelos usados podem ser observados nas tabelas presentes no Apêndice A. O melhor resultado obtido nesses 1000 experimentos, mostrado na Tabela 5.5 foi gerado com o modelo *Logistic Regression*, para a *feature* F_{mm_d} e com a combinação de filtros $4v\_5p$, conseguindo uma acurácia média de 60,1%.

O resultado se mostrou pior do que o encontrado sem a utilização desses filtros, de forma a eliminar uma das possíveis soluções pensadas para o problema. Também foi possível observar nos resultados no Apêndice A que não houve relação direta entre o aumento das pausas das combinações e a acurácia obtida, dado que houveram diversos casos em que para uma mesma *feature* e modelo, combinações com quantidade de pausas menores alcançaram acurácias maiores.

Esse resultado evidencia também a possibilidade de que as *features* não sejam tão representativas quanto previamente pensado para o problema. Contudo, é importante notar que há formas de se melhorar o resultado, seja com novas *features*, com novos parâmetros para treinamento dos classificadores ou até mesmo a utilização de novos classificadores.

6 Conclusões e Trabalhos Futuros

O trabalho teve como objetivo a identificação automática das pausas de leitura e o estudo de abordagens capazes de classificar automaticamente o respeito às pausas durante a leitura. Foram apresentados os conceitos relacionados à avaliação da fluência, prosódia e seus componentes, sendo a pausa um deles, reforçando a importância do trabalho para a realização de uma avaliação da leitura mais completa.

Para a realização do trabalho, foram selecionados áudios de 1 minuto de leitura em voz alta de crianças em fase de alfabetização, dividindo-os em 79 áudios para base de investigação e 859 para base de teste. A base de investigação foi utilizada para definir as melhores métricas e parâmetros dos modelos de classificação para o teste com a base maior de teste. Os áudios foram classificados manualmente por avaliadores treinados pela Fundação CAEd/UFJF, a qual forneceu os áudios para realização dos experimentos. Uma das informações obtidas através da avaliação manual de um determinado áudio é quanto ao seu respeito às pausas de sentido durante a leitura toda, indicando se as pausas foram respeitadas ou não.

Durante os experimentos, foram criadas 8 *features* para representar os áudios e serem utilizadas no treinamento de modelos de aprendizado de máquinas para classificar automaticamente os áudios quanto ao respeito às pausas. Todas as *features* utilizavam dados das pausas e seus tipos, separados entre as normais e as de vírgula e de ponto, as quais se esperavam durações maiores que o resto das pausas durante a leitura. A *feature* que obteve o melhor resultado para o melhor modelo foi a F_{mm_r}, cujas informações passadas são a razão entre a mediana das pausas de vírgula e de ponto com a mediana das pausas normais.

Os modelos treinados para classificação tiveram os principais parâmetros testados com diferentes valores, apesar de ainda terem muitos parâmetros possíveis de serem verificados devido ao tempo demandado para os testes. Uma tarefa futura é testar mais variações de parâmetros para treinar o modelo de forma a encontrar a combinação mais adequada para esse tipo e quantidade de dados. Os modelos utilizados consistiram no *Ran-*

dom Forest Classifier, *Logistic Regression*, *Voting Classifier*, *XGBoost* e *MLP Classifier*, o último sendo o que obteve o melhor resultado dentre os testados, com **69,5%** $\pm 3,4\%$ de acurácia.

Apesar da identificação das pausas ter sido correta, a sua classificação automática não obteve resultado igualmente satisfatório, tendo obtido acurácia máxima menor que a média obtida no melhor resultado da base de investigação. Em um primeiro momento, esperava-se que uma base maior gerasse o treinamento de modelos mais capazes de classificar adequadamente os áudios.

Ocorre que para a base de investigação, apenas os melhores áudios foram selecionados, sem um único erro durante a leitura, sobrando para a base de teste apenas áudios com ocorrência de pelo menos um erro durante a leitura. Para evitar alinhamentos de pausas errados devido a palavras lidas erradas em torno da pausa, foram consideradas apenas pausas que ocorreram entre duas leituras de palavras corretas. O problema é que a leitura de uma criança tende a ser lento e pode conter quantidades consideráveis de erros durante a leitura, o que faz com que haja poucas ocorrências de pausas por vírgula e ponto durante a leitura em 1 minuto de áudio, as quais já são poucas durante o texto completo a ser lido (em torno de 6 vírgulas e 9 pontos).

As pausas de vírgula e ponto que já ocorrem poucas vezes durante a leitura das crianças, são diminuídas mais ainda devido ao filtro de considerar apenas pausas entre palavras lidas corretamente, de forma que se a vírgula ocorre antes de uma palavra lida incorretamente ela será desconsiderada por exemplo. Foi possível verificar, ao inspecionar a execução do algoritmo que seleciona as pausas, 356 casos em que o resultado final não havia pausa alguma de vírgula e/ou ponto, equivalendo a 41,5% da base de teste. A falta de pausas de qualquer tipo podem fazer com que a maioria das *features* utilizadas fiquem com valores prejudicados, o que atrapalha o treinamento do modelo de forma adequada, dada que as *features* possivelmente não representam adequadamente os áudios nessas situações. Então considerando a quantidade de suas ocorrências, o modelo pode ter sido impactado negativamente.

Foi testada uma solução que consistia na filtragem dos áudios utilizados de forma a desconsiderar áudios que não cumprissem um determinado limiar definido em relação à

presença de pausas de vírgula e de ponto nos áudios. Os limiares variaram entre ter de 1 a 5 pausas de vírgula no mínimo, da mesma forma que para as pausas de ponto. Detalhes de sua execução podem ser encontradas na seção 5.2 e no Apêndice A. O melhor resultado obtido dentre todos os filtros alcançou apenas $60,1\% \pm 9,5\%$ de acurácia. O filtro com melhor resultado deixou na base 218 áudios a serem considerados, em que 121 deles são de leituras que respeitaram as pausas e 97 que desrespeitaram. A acurácia obtida através dos filtros foi menor que o alcançado sem sua utilização, demonstrando que a solução foi ineficaz para o problema.

Portanto, como próximos passos, foram pensadas em algumas formas para melhorar os resultados. Uma possibilidade é que a de que as *features* definidas não sejam adequadas para representar os áudios no problema proposto, a definição de *features* mais adequadas se fazem necessárias então. Também há formas de melhorar os modelos utilizados ou até mesmo incluir novos modelos. Para melhorar os modelos, a ideia é testar diferentes valores para os parâmetros de cada classificador, de forma a encontrar os que gerem o melhor resultado para a base teste.

Bibliografia

ANANTHAKRISHNAN, S.; NARAYANAN, S. Unsupervised adaptation of categorical prosody models for prosody labeling and speech recognition. *IEEE transactions on audio, speech, and language processing*, IEEE, v. 17, n. 1, p. 138–149, 2009.

ARANTES, P. Implementação em praat de algoritmos para descrição de correlatos acústicos da prosódia da fala. *II JORNADA DE DESCRIçãO DO PORTUgUêS, II*, p. 32–38, 2011.

BAJO, M. D.; FARRÚS, M.; WANNER, L. An automatic prosody tagger for spontaneous speech. In: COLING. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers; 2016 Dec 11-17; Osaka, Japan.[Unknonw place]:[COLING]; 2016. p. 377-87.* [S.l.], 2016.

BARBOSA, P. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. *REVISTA DE ESTUDOS DA LINGUAGEM*, v. 20, n. 1, p. 11–27, 2012. ISSN 2237-2083. Disponível em: ⟨http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2571⟩.

BARROS, M. T. d. A. A relação entre compreensão leitora e prosódia em crianças. Universidade Federal de Pernambuco, 2017.

BARROS, R. C. R. de. A natureza lingüística da alfabetização: aspectos prosódicos. *Signótica*, v. 6, n. 1, p. 119–130, 1994.

BENJAMIN, R. G.; SCHWANENFLUGEL, P. J. Text complexity and oral reading prosody in young readers. *Reading Research Quarterly*, Wiley Online Library, v. 45, n. 4, p. 388–404, 2010.

BENJAMIN, R. G.; SCHWANENFLUGEL, P. J.; MEISINGER, E. B.; GROFF, C.; KUHN, M. R.; STEINER, L. A spectrographically grounded scale for evaluating reading expressiveness. *Reading Research Quarterly*, Wiley Online Library, v. 48, n. 2, p. 105–133, 2013.

BERGNER, Y.; DAVIER, A. A. von. Process data in naep: Past, present, and future. *Journal of Educational and Behavioral Statistics*, SAGE Publications Sage CA: Los Angeles, CA, v. 44, n. 6, p. 706–732, 2019.

BOLAÑOS, D.; COLE, R. A.; WARD, W. H.; TINDAL, G. A.; SCHWANENFLUGEL, P. J.; KUHN, M. R. Automatic assessment of expressive oral reading. *Speech Communication*, Elsevier, v. 55, n. 2, p. 221–236, 2013.

BOLAÑOS, D.; COLE, R. A.; WARD, W. H.; TINDAL, G. A.; HASBROUCK, J.; SCHWANENFLUGEL, P. J. Human and automated assessment of oral reading fluency. *Journal of educational psychology*, American Psychological Association, v. 105, n. 4, p. 1142, 2013.

BRAGA, J. N.; OLIVEIRA, D. S. F. d.; SAMPAIO, T. M. M. Frequência fundamental da voz de crianças. *Revista CEFAC*, SciELO Brasil, v. 11, p. 119–126, 2009.

CARCHEDI, L. C.; BARRÉRE, E.; SOUZA, J. F. de. Avalia online: um sistema para avaliação em larga escala de testes de fluência de leitura. In: SBC. *Anais do XXXII Simpósio Brasileiro de Informática na Educação*. [S.l.], 2021. p. 01–11.

CONNELLY, L. Logistic regression. *Medsurg Nursing*, Anthony J. Jannetti, Inc., v. 29, n. 5, p. 353–354, 2020.

DUONG, M.; MOSTOW, J.; SITARAM, S. Two methods for assessing oral reading prosody. *ACM Transactions on Speech and Language Processing (TSLP)*, ACM New York, NY, USA, v. 7, n. 4, p. 1–22, 2011.

EL-KENAWY, E.-S. M.; IBRAHIM, A.; MIRJALILI, S.; EID, M. M.; HUSSEIN, S. E. Novel feature selection and voting classifier algorithms for covid-19 classification in ct images. *IEEE access*, IEEE, v. 8, p. 179317–179335, 2020.

FERREIRA, R. D. S. *Avaliação da fluência na leitura em crianças com e sem necessidades educativas especiais: Validação de uma prova de fluência na leitura para o 2º Ano do 1º CEB*. Tese (Doutorado), 2009.

FILIPE, M. G.; VICENTE, S. Avaliação da competência prosódica de segmentação em crianças e adultos. *Actas do VII simpósio nacional de investigação em psicologia*, 2010.

GODDE, E.; BAILLY, G.; BOSSE, M.-L. Reading prosody development: automatic assessment for a longitudinal study. In: *SLaTE 2019-8th ISCA Workshop on Speech and Language Technology in Education*. [S.l.: s.n.], 2019.

JR, R. S.; KAFKA, S.; SEARA, I.; PACHECO, F.; KLEIN, S.; SEARA, R. Parâmetros lingüísticos utilizados para a geração automática de prosódia em sistemas de síntese de fala. *XXI simpósio brasileiro de telecomunicações*, p. 1–6, 2004.

JUNIOR, A. C.; CASANOVA, E.; SOARES, A.; OLIVEIRA, F. S. de; OLIVEIRA, L.; JUNIOR, R. C. F.; SILVA, D. P. P. da; FAYET, F. G.; CARLOTTO, B. B.; GRIS, L. R. S. et al. Coraa: a large corpus of spontaneous and prepared speech manually validated for speech recognition in brazilian portuguese. *arXiv preprint arXiv:2110.15731*, 2021.

KAMEL, H.; ABDULAH, D.; AL-TUWAIJARI, J. M. Cancer classification using gaussian naive bayes algorithm. In: IEEE. *2019 International Engineering Conference (IEC)*. [S.l.], 2019. p. 165–170.

LOPES, J.; SILVA, M. M.; MONIZ, A.; SPEAR-SWERLING, L.; ZIBULSKY, J. Evolução da prosódia e compreensão da leitura: Um estudo longitudinal do 2. º ano ao final do 3. º ano de escolaridade. *Revista de Psicodidáctica*, Universidad del País Vasco/Euskal Herriko Unibertsitatea, v. 20, n. 1, p. 5–23, 2015.

PAL, M. Random forest classifier for remote sensing classification. *International journal of remote sensing*, Taylor & Francis, v. 26, n. 1, p. 217–222, 2005.

PARDINHO, V. da M. A relevância da prosódia no processo de alfabetização/letramento. In: *Caderno de Resumos do Congresso de Leitura do Brasil*. [S.l.: s.n.], 2021. v. 1, n. 1.

PINTO, J. C. B. R.; NAVAS, A. L. G. P. Efeitos da estimulação da fluência de leitura com ênfase na prosódia. *Jornal da Sociedade Brasileira de Fonoaudiologia*, SciELO Brasil, v. 23, p. 21–26, 2011.

PROSóDIA. 2018. Disponível em: ⟨https://www.dicio.com.br/prosodia/⟩.

PULIEZI, S.; MALUF, M. R. A fluência e sua importância para a compreensão da leitura. *Psico-USF*, SciELO Brasil, v. 19, p. 467–475, 2014.

QIAN, Y.; WU, Z.; MA, X.; SOONG, F. Automatic prosody prediction and detection with conditional random field (crf) models. In: IEEE. *2010 7th International Symposium on Chinese Spoken Language Processing*. [S.l.], 2010. p. 135–138.

QIU, Y.; ZHOU, J.; KHANDELWAL, M.; YANG, H.; YANG, P.; LI, C. Performance evaluation of hybrid woa-xgboost, gwo-xgboost and bo-xgboost models to predict blast-induced ground vibration. *Engineering with Computers*, Springer, v. 38, n. 5, p. 4145–4162, 2022.

RASINSKI, T. V. Assessing reading fluency. *Pacific Resources for Education and Learning (PREL)*, ERIC, 2004.

SABU, K.; RAO, P. Prosodic event detection in children's read speech. *Computer Speech & Language*, Elsevier, v. 68, p. 101200, 2021.

SABU, K.; SWARUP, P.; TULSIANI, H.; RAO, P. Automatic assessment of children's l2 reading for accuracy and fluency. In: *SLaTE*. [S.l.: s.n.], 2017. p. 121–126.

SCHWANENFLUGEL, P. J.; BENJAMIN, R. G. Lexical prosody as an aspect of oral reading fluency. *Reading and Writing*, Springer, v. 30, n. 1, p. 143–162, 2017.

SHAHNAWAZUDDIN, S.; ADIGA, N.; KATHANIA, H. K.; SAI, B. T. Creating speaker independent asr system through prosody modification based data augmentation. *Pattern Recognition Letters*, Elsevier, v. 131, p. 213–218, 2020.

SKERRY-RYAN, R.; BATTENBERG, E.; XIAO, Y.; WANG, Y.; STANTON, D.; SHOR, J.; WEISS, R.; CLARK, R.; SAUROUS, R. A. Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. In: DY, J.; KRAUSE, A. (Ed.). *Proceedings of the 35th International Conference on Machine Learning*. PMLR, 2018. (Proceedings of Machine Learning Research, v. 80), p. 4693–4702. Disponível em: ⟨https://proceedings.mlr.press/v80/skerry-ryan18a.html⟩.

SOUSA, S. Z. Concepções de qualidade da educação básica forjadas por meio de avaliações em larga escala. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, SciELO Brasil, v. 19, p. 407–420, 2014.

SRIDHAR, V. K. R.; BANGALORE, S.; NARAYANAN, S. S. Exploiting acoustic and syntactic features for automatic prosody labeling in a maximum entropy framework. *IEEE transactions on audio, speech, and language processing*, IEEE, v. 16, n. 4, p. 797–811, 2008.

TEIXEIRA, B.; BARBOSA, P. A.; RASO, T. Para a segmentação automática de fronteira na fala espontânea a partir de parâmetros prosódicos. *Linguística de Corpus*, p. 425.

TEIXEIRA, B. H. F. Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea. Universidade Federal de Minas Gerais, 2018.

VEENENDAAL, N. J.; GROEN, M. A.; VERHOEVEN, L. What oral text reading fluency can reveal about reading comprehension. *Journal of Research in Reading*, Wiley Online Library, v. 38, n. 3, p. 213–225, 2015.

WINDEATT, T. Ensemble mlp classifier design. In: *Computational Intelligence Paradigms*. [S.l.]: Springer, 2008. p. 133–147.

YANG, M.; HIRSCHI, K.; LOONEY, S. D.; KANG, O.; HANSEN, J. H. Improving mispronunciation detection with wav2vec2-based momentum pseudo-labeling for accentedness and intelligibility assessment. *arXiv preprint arXiv:2203.15937*, 2022.

ZHANG, H.; SONG, Y.; SONG, H.-T. Construction of ontology-based user model for web personalization. In: CONATI, C.; MCCOY, K. F.; PALIOURAS, G. (Ed.). *Proceedings of the 11 International Conference on User Modeling*. Corfu, Greece: Springer, 2007. (Lecture Notes in Computer Science, v. 4511), p. 67–76. ISBN 978-3-540-73077-4.

A - Resultados de experimentos em cada modelo

Tabela A.1: Acurácia média (ACC) resultante dos experimentos para o *Random Forest Classifier*, para cada combinação (Comb.) e *feature*

Comb.	***feature***	**ACC**	**Comb.**	***feature***	**ACC**	**Comb.**	***feature***	**ACC**
1v_1p	F_{mm}	53.3%	2v_4p	F_{mm_r}	52.8%	4v_2p	$F_{r\_q2vp\_q3n}$	54.8%
1v_1p	F_{mm_d}	52.5%	2v_4p	$F_{d\_q2vp\_q3n}$	47.3%	4v_2p	F_{all_d}	49.4%
1v_1p	F_{mm_r}	53.7%	2v_4p	$F_{r\_q2vp\_q3n}$	54.0%	4v_2p	F_{all_r}	46.6%
1v_1p	$F_{d\_q2vp\_q3n}$	47.3%	2v_4p	F_{all_d}	49.4%	4v_2p	F_{all}	49.9%
1v_1p	$F_{r\_q2vp\_q3n}$	50.5%	2v_4p	F_{all_r}	51.8%	4v_3p	F_{mm}	53.0%
1v_1p	F_{all_d}	50.7%	2v_4p	F_{all}	51.3%	4v_3p	F_{mm_d}	47.6%
1v_1p	F_{all_r}	51.1%	2v_5p	F_{mm}	52.6%	4v_3p	F_{mm_r}	44.8%
1v_1p	F_{all}	53.5%	2v_5p	F_{mm_d}	49.6%	4v_3p	$F_{d\_q2vp\_q3n}$	51.6%
1v_2p	F_{mm}	52.0%	2v_5p	F_{mm_r}	52.1%	4v_3p	$F_{r\_q2vp\_q3n}$	54.8%
1v_2p	F_{mm_d}	50.9%	2v_5p	$F_{d\_q2vp\_q3n}$	47.4%	4v_3p	F_{all_d}	49.4%
1v_2p	F_{mm_r}	52.2%	2v_5p	$F_{r\_q2vp\_q3n}$	53.3%	4v_3p	F_{all_r}	49.8%
1v_2p	$F_{d\_q2vp\_q3n}$	49.3%	2v_5p	F_{all_d}	50.8%	4v_3p	F_{all}	49.0%
1v_2p	$F_{r\_q2vp\_q3n}$	55.1%	2v_5p	F_{all_r}	50.1%	4v_4p	F_{mm}	54.0%
1v_2p	F_{all_d}	48.3%	2v_5p	F_{all}	52.3%	4v_4p	F_{mm_d}	49.4%
1v_2p	F_{all_r}	52.2%	3v_1p	F_{mm}	50.3%	4v_4p	F_{mm_r}	47.6%
1v_2p	F_{all}	50.2%	3v_1p	F_{mm_d}	48.9%	4v_4p	$F_{d\_q2vp\_q3n}$	52.9%
1v_3p	F_{mm}	49.5%	3v_1p	F_{mm_r}	51.7%	4v_4p	$F_{r\_q2vp\_q3n}$	55.8%
1v_3p	F_{mm_d}	50.1%	3v_1p	$F_{d\_q2vp\_q3n}$	47.8%	4v_4p	F_{all_d}	47.6%
1v_3p	F_{mm_r}	51.4%	3v_1p	$F_{r\_q2vp\_q3n}$	52.2%	4v_4p	F_{all_r}	47.1%
1v_3p	$F_{d\_q2vp\_q3n}$	48.4%	3v_1p	F_{all_d}	52.0%	4v_4p	F_{all}	49.4%
1v_3p	$F_{r\_q2vp\_q3n}$	52.9%	3v_1p	F_{all_r}	51.1%	4v_5p	F_{mm}	54.7%
1v_3p	F_{all_d}	49.9%	3v_1p	F_{all}	53.9%	4v_5p	F_{mm_d}	46.4%
1v_3p	F_{all_r}	48.2%	3v_2p	F_{mm}	50.0%	4v_5p	F_{mm_r}	47.3%
1v_3p	F_{all}	53.5%	3v_2p	F_{mm_d}	48.6%	4v_5p	$F_{d\_q2vp\_q3n}$	53.2%
1v_4p	F_{mm}	54.5%	3v_2p	F_{mm_r}	52.2%	4v_5p	$F_{r\_q2vp\_q3n}$	56.0%
1v_4p	F_{mm_d}	49.4%	3v_2p	$F_{d\_q2vp\_q3n}$	48.1%	4v_5p	F_{all_d}	51.0%
1v_4p	F_{mm_r}	51.7%	3v_2p	$F_{r\_q2vp\_q3n}$	53.3%	4v_5p	F_{all_r}	52.8%
1v_4p	$F_{d\_q2vp\_q3n}$	47.0%	3v_2p	F_{all_d}	51.7%	4v_5p	F_{all}	54.7%
1v_4p	$F_{r\_q2vp\_q3n}$	53.7%	3v_2p	F_{all_r}	47.8%	5v_1p	F_{mm}	49.9%
1v_4p	F_{all_d}	50.2%	3v_2p	F_{all}	48.9%	5v_1p	F_{mm_d}	39.3%
1v_4p	F_{all_r}	51.7%	3v_3p	F_{mm}	49.5%	5v_1p	F_{mm_r}	47.9%
1v_4p	F_{all}	53.5%	3v_3p	F_{mm_d}	47.8%	5v_1p	$F_{d\_q2vp\_q3n}$	42.6%
1v_5p	F_{mm}	53.3%	3v_3p	F_{mm_r}	52.2%	5v_1p	$F_{r\_q2vp\_q3n}$	55.2%
1v_5p	F_{mm_d}	52.6%	3v_3p	$F_{d\_q2vp\_q3n}$	49.2%	5v_1p	F_{all_d}	42.5%

Continua na próxima página

Tabela A.1 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_5p	F_{mm_r}	50.4%	3v_3p	$F_{r\_q2vp\_q3n}$	53.0%	5v_1p	F_{all_r}	52.1%
1v_5p	$F_{d\_q2vp\_q3n}$	47.6%	3v_3p	F_{all_d}	50.9%	5v_1p	F_{all}	43.6%
1v_5p	$F_{r\_q2vp\_q3n}$	53.0%	3v_3p	F_{all_r}	48.0%	5v_2p	F_{mm}	50.9%
1v_5p	F_{all_d}	50.4%	3v_3p	F_{all}	50.5%	5v_2p	F_{mm_d}	43.6%
1v_5p	F_{all_r}	50.4%	3v_4p	F_{mm}	46.9%	5v_2p	F_{mm_r}	47.9%
1v_5p	F_{all}	52.6%	3v_4p	F_{mm_d}	51.4%	5v_2p	$F_{d\_q2vp\_q3n}$	42.6%
2v_1p	F_{mm}	51.2%	3v_4p	F_{mm_r}	52.8%	5v_2p	$F_{r\_q2vp\_q3n}$	55.2%
2v_1p	F_{mm_d}	50.0%	3v_4p	$F_{d\_q2vp\_q3n}$	46.4%	5v_2p	F_{all_d}	43.5%
2v_1p	F_{mm_r}	53.3%	3v_4p	$F_{r\_q2vp\_q3n}$	53.1%	5v_2p	F_{all_r}	51.0%
2v_1p	$F_{d\_q2vp\_q3n}$	49.1%	3v_4p	F_{all_d}	49.5%	5v_2p	F_{all}	50.0%
2v_1p	$F_{r\_q2vp\_q3n}$	54.2%	3v_4p	F_{all_r}	49.2%	5v_3p	F_{mm}	51.0%
2v_1p	F_{all_d}	48.1%	3v_4p	F_{all}	48.3%	5v_3p	F_{mm_d}	42.5%
2v_1p	F_{all_r}	50.0%	3v_5p	F_{mm}	49.7%	5v_3p	F_{mm_r}	45.8%
2v_1p	F_{all}	51.2%	3v_5p	F_{mm_d}	51.1%	5v_3p	$F_{d\_q2vp\_q3n}$	42.5%
2v_2p	F_{mm}	54.1%	3v_5p	F_{mm_r}	50.1%	5v_3p	$F_{r\_q2vp\_q3n}$	57.4%
2v_2p	F_{mm_d}	51.8%	3v_5p	$F_{d\_q2vp\_q3n}$	46.1%	5v_3p	F_{all_d}	44.6%
2v_2p	F_{mm_r}	55.6%	3v_5p	$F_{r\_q2vp\_q3n}$	50.6%	5v_3p	F_{all_r}	47.8%
2v_2p	$F_{d\_q2vp\_q3n}$	47.3%	3v_5p	F_{all_d}	48.6%	5v_3p	F_{all}	46.7%
2v_2p	$F_{r\_q2vp\_q3n}$	54.1%	3v_5p	F_{all_r}	48.3%	5v_4p	F_{mm}	51.0%
2v_2p	F_{all_d}	50.2%	3v_5p	F_{all}	48.6%	5v_4p	F_{mm_d}	43.6%
2v_2p	F_{all_r}	52.2%	4v_1p	F_{mm}	50.3%	5v_4p	F_{mm_r}	45.8%
2v_2p	F_{all}	50.8%	4v_1p	F_{mm_d}	49.8%	5v_4p	$F_{d\_q2vp\_q3n}$	43.6%
2v_3p	F_{mm}	54.4%	4v_1p	F_{mm_r}	46.6%	5v_4p	$F_{r\_q2vp\_q3n}$	56.3%
2v_3p	F_{mm_d}	50.6%	4v_1p	$F_{d\_q2vp\_q3n}$	52.4%	5v_4p	F_{all_d}	43.5%
2v_3p	F_{mm_r}	50.1%	4v_1p	$F_{r\_q2vp\_q3n}$	53.4%	5v_4p	F_{all_r}	50.0%
2v_3p	$F_{d\_q2vp\_q3n}$	46.6%	4v_1p	F_{all_d}	49.4%	5v_4p	F_{all}	47.9%
2v_3p	$F_{r\_q2vp\_q3n}$	54.4%	4v_1p	F_{all_r}	48.4%	5v_5p	F_{mm}	53.2%
2v_3p	F_{all_d}	50.4%	4v_1p	F_{all}	49.4%	5v_5p	F_{mm_d}	43.6%
2v_3p	F_{all_r}	50.2%	4v_2p	F_{mm}	50.8%	5v_5p	F_{mm_r}	45.8%
2v_3p	F_{all}	49.9%	4v_2p	F_{mm_d}	48.5%	5v_5p	$F_{d\_q2vp\_q3n}$	42.6%
2v_4p	F_{mm}	50.8%	4v_2p	F_{mm_r}	46.2%	5v_5p	$F_{r\_q2vp\_q3n}$	53.2%
2v_4p	F_{mm_d}	50.3%	4v_2p	$F_{d\_q2vp\_q3n}$	52.5%	5v_5p	F_{all_d}	44.6%
						5v_5p	F_{all_r}	52.1%
						5v_5p	F_{all}	47.9%

Tabela A.2: Acurácia média (ACC) resultante dos experimentos para o *Logistic Regression*, para cada combinação (Comb.) e *feature*

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_1p	F_{mm}	55.7%	2v_4p	F_{mm_r}	53.8%	4v_2p	$F_{r\_q2vp\_q3n}$	54.8%
1v_1p	F_{mm_d}	51.7%	2v_4p	$F_{d\_q2vp\_q3n}$	53.2%	4v_2p	F_{all_d}	56.7%

Continua na próxima página

Tabela A.2 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_1p	F_{mm_r}	53.5%	2v_4p	$F_{r\_q2vp\_q3n}$	54.0%	4v_2p	F_{all_r}	51.7%
1v_1p	$F_{d\_q2vp\_q3n}$	52.7%	2v_4p	F_{all_d}	54.9%	4v_2p	F_{all}	56.7%
1v_1p	$F_{r\_q2vp\_q3n}$	53.1%	2v_4p	F_{all_r}	53.2%	4v_3p	F_{mm}	57.6%
1v_1p	F_{all_d}	53.7%	2v_4p	F_{all}	54.7%	4v_3p	F_{mm_d}	58.1%
1v_1p	F_{all_r}	54.1%	2v_5p	F_{mm}	53.0%	4v_3p	F_{mm_r}	52.1%
1v_1p	F_{all}	53.9%	2v_5p	F_{mm_d}	50.4%	4v_3p	$F_{d\_q2vp\_q3n}$	58.5%
1v_2p	F_{mm}	54.4%	2v_5p	F_{mm_r}	50.6%	4v_3p	$F_{r\_q2vp\_q3n}$	54.8%
1v_2p	F_{mm_d}	51.8%	2v_5p	$F_{d\_q2vp\_q3n}$	53.8%	4v_3p	F_{all_d}	56.7%
1v_2p	F_{mm_r}	54.7%	2v_5p	$F_{r\_q2vp\_q3n}$	54.5%	4v_3p	F_{all_r}	51.7%
1v_2p	$F_{d\_q2vp\_q3n}$	52.6%	2v_5p	F_{all_d}	52.8%	4v_3p	F_{all}	56.7%
1v_2p	$F_{r\_q2vp\_q3n}$	53.4%	2v_5p	F_{all_r}	53.5%	4v_4p	F_{mm}	57.6%
1v_2p	F_{all_d}	54.0%	2v_5p	F_{all}	52.3%	4v_4p	F_{mm_d}	58.1%
1v_2p	F_{all_r}	54.0%	3v_1p	F_{mm}	55.3%	4v_4p	F_{mm_r}	52.1%
1v_2p	F_{all}	54.0%	3v_1p	F_{mm_d}	53.6%	4v_4p	$F_{d\_q2vp\_q3n}$	58.5%
1v_3p	F_{mm}	54.6%	3v_1p	F_{mm_r}	55.3%	4v_4p	$F_{r\_q2vp\_q3n}$	54.8%
1v_3p	F_{mm_d}	52.9%	3v_1p	$F_{d\_q2vp\_q3n}$	55.0%	4v_4p	F_{all_d}	56.7%
1v_3p	F_{mm_r}	53.7%	3v_1p	$F_{r\_q2vp\_q3n}$	55.3%	4v_4p	F_{all_r}	51.7%
1v_3p	$F_{d\_q2vp\_q3n}$	54.3%	3v_1p	F_{all_d}	54.5%	4v_4p	F_{all}	56.7%
1v_3p	$F_{r\_q2vp\_q3n}$	53.5%	3v_1p	F_{all_r}	53.4%	4v_5p	F_{mm}	56.9%
1v_3p	F_{all_d}	54.2%	3v_1p	F_{all}	54.5%	4v_5p	F_{mm_d}	60.1%
1v_3p	F_{all_r}	53.7%	3v_2p	F_{mm}	55.3%	4v_5p	F_{mm_r}	52.4%
1v_3p	F_{all}	54.4%	3v_2p	F_{mm_d}	53.6%	4v_5p	$F_{d\_q2vp\_q3n}$	59.2%
1v_4p	F_{mm}	56.2%	3v_2p	F_{mm_r}	55.3%	4v_5p	$F_{r\_q2vp\_q3n}$	53.8%
1v_4p	F_{mm_d}	52.1%	3v_2p	$F_{d\_q2vp\_q3n}$	55.0%	4v_5p	F_{all_d}	57.4%
1v_4p	F_{mm_r}	54.3%	3v_2p	$F_{r\_q2vp\_q3n}$	55.3%	4v_5p	F_{all_r}	51.9%
1v_4p	$F_{d\_q2vp\_q3n}$	54.1%	3v_2p	F_{all_d}	54.5%	4v_5p	F_{all}	56.9%
1v_4p	$F_{r\_q2vp\_q3n}$	54.5%	3v_2p	F_{all_r}	53.4%	5v_1p	F_{mm}	54.3%
1v_4p	F_{all_d}	53.9%	3v_2p	F_{all}	54.5%	5v_1p	F_{mm_d}	55.4%
1v_4p	F_{all_r}	52.1%	3v_3p	F_{mm}	55.3%	5v_1p	F_{mm_r}	53.2%
1v_4p	F_{all}	53.4%	3v_3p	F_{mm_d}	53.6%	5v_1p	$F_{d\_q2vp\_q3n}$	44.6%
1v_5p	F_{mm}	55.4%	3v_3p	F_{mm_r}	55.3%	5v_1p	$F_{r\_q2vp\_q3n}$	47.9%
1v_5p	F_{mm_d}	54.1%	3v_3p	$F_{d\_q2vp\_q3n}$	55.0%	5v_1p	F_{all_d}	53.2%
1v_5p	F_{mm_r}	53.7%	3v_3p	$F_{r\_q2vp\_q3n}$	55.3%	5v_1p	F_{all_r}	50.0%
1v_5p	$F_{d\_q2vp\_q3n}$	54.6%	3v_3p	F_{all_d}	54.5%	5v_1p	F_{all}	53.2%
1v_5p	$F_{r\_q2vp\_q3n}$	54.6%	3v_3p	F_{all_r}	53.4%	5v_2p	F_{mm}	54.3%
1v_5p	F_{all_d}	55.0%	3v_3p	F_{all}	54.5%	5v_2p	F_{mm_d}	55.4%
1v_5p	F_{all_r}	53.0%	3v_4p	F_{mm}	53.1%	5v_2p	F_{mm_r}	53.2%
1v_5p	F_{all}	54.6%	3v_4p	F_{mm_d}	54.7%	5v_2p	$F_{d\_q2vp\_q3n}$	44.6%
2v_1p	F_{mm}	53.8%	3v_4p	F_{mm_r}	55.6%	5v_2p	$F_{r\_q2vp\_q3n}$	47.9%
2v_1p	F_{mm_d}	51.9%	3v_4p	$F_{d\_q2vp\_q3n}$	55.1%	5v_2p	F_{all_d}	53.2%
2v_1p	F_{mm_r}	54.7%	3v_4p	$F_{r\_q2vp\_q3n}$	55.9%	5v_2p	F_{all_r}	50.0%
2v_1p	$F_{d\_q2vp\_q3n}$	53.1%	3v_4p	F_{all_d}	54.5%	5v_2p	F_{all}	53.2%
2v_1p	$F_{r\_q2vp\_q3n}$	55.0%	3v_4p	F_{all_r}	55.3%	5v_3p	F_{mm}	54.3%

Continua na próxima página

Tabela A.2 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
2v_1p	F_{all_d}	55.4%	3v_4p	F_{all}	55.3%	5v_3p	F_{mm_d}	55.4%
2v_1p	F_{all_r}	55.0%	3v_5p	F_{mm}	54.0%	5v_3p	F_{mm_r}	53.2%
2v_1p	F_{all}	55.2%	3v_5p	F_{mm_d}	54.2%	5v_3p	$F_{d\_q2vp\_q3n}$	44.6%
2v_2p	F_{mm}	53.7%	3v_5p	F_{mm_r}	55.1%	5v_3p	$F_{r\_q2vp\_q3n}$	47.9%
2v_2p	F_{mm_d}	52.0%	3v_5p	$F_{d\_q2vp\_q3n}$	56.0%	5v_3p	F_{all_d}	53.2%
2v_2p	F_{mm_r}	53.4%	3v_5p	$F_{r\_q2vp\_q3n}$	54.8%	5v_3p	F_{all_r}	50.0%
2v_2p	$F_{d\_q2vp\_q3n}$	53.6%	3v_5p	F_{all_d}	54.3%	5v_3p	F_{all}	53.2%
2v_2p	$F_{r\_q2vp\_q3n}$	55.3%	3v_5p	F_{all_r}	56.5%	5v_4p	F_{mm}	54.3%
2v_2p	F_{all_d}	55.8%	3v_5p	F_{all}	54.8%	5v_4p	F_{mm_d}	55.4%
2v_2p	F_{all_r}	55.1%	4v_1p	F_{mm}	57.6%	5v_4p	F_{mm_r}	53.2%
2v_2p	F_{all}	55.6%	4v_1p	F_{mm_d}	58.1%	5v_4p	$F_{d\_q2vp\_q3n}$	44.6%
2v_3p	F_{mm}	52.3%	4v_1p	F_{mm_r}	52.1%	5v_4p	$F_{r\_q2vp\_q3n}$	47.9%
2v_3p	F_{mm_d}	49.9%	4v_1p	$F_{d\_q2vp\_q3n}$	58.5%	5v_4p	F_{all_d}	53.2%
2v_3p	F_{mm_r}	54.6%	4v_1p	$F_{r\_q2vp\_q3n}$	54.8%	5v_4p	F_{all_r}	50.0%
2v_3p	$F_{d\_q2vp\_q3n}$	52.5%	4v_1p	F_{all_d}	56.7%	5v_4p	F_{all}	53.2%
2v_3p	$F_{r\_q2vp\_q3n}$	54.6%	4v_1p	F_{all_r}	51.7%	5v_5p	F_{mm}	54.3%
2v_3p	F_{all_d}	54.2%	4v_1p	F_{all}	56.7%	5v_5p	F_{mm_d}	55.4%
2v_3p	F_{all_r}	52.5%	4v_2p	F_{mm}	57.6%	5v_5p	F_{mm_r}	53.2%
2v_3p	F_{all}	53.9%	4v_2p	F_{mm_d}	58.1%	5v_5p	$F_{d\_q2vp\_q3n}$	44.6%
2v_4p	F_{mm}	54.2%	4v_2p	F_{mm_r}	52.1%	5v_5p	$F_{r\_q2vp\_q3n}$	47.9%
2v_4p	F_{mm_d}	50.8%	4v_2p	$F_{d\_q2vp\_q3n}$	58.5%	5v_5p	F_{all_d}	53.2%
						5v_5p	F_{all_r}	50.0%
						5v_5p	F_{all}	53.2%

Tabela A.3: Acurácia média (ACC) resultante dos experimentos para o *Voting Classifier*, para cada combinação (Comb.) e *feature*

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_1p	F_{mm}	54.5%	2v_4p	F_{mm_r}	57.1%	4v_2p	$F_{r\_q2vp\_q3n}$	53.9%
1v_1p	F_{mm_d}	52.7%	2v_4p	$F_{d\_q2vp\_q3n}$	47.7%	4v_2p	F_{all_d}	53.9%
1v_1p	F_{mm_r}	56.7%	2v_4p	$F_{r\_q2vp\_q3n}$	55.9%	4v_2p	F_{all_r}	49.3%
1v_1p	$F_{d\_q2vp\_q3n}$	48.5%	2v_4p	F_{all_d}	56.1%	4v_2p	F_{all}	53.0%
1v_1p	$F_{r\_q2vp\_q3n}$	53.0%	2v_4p	F_{all_r}	54.9%	4v_3p	F_{mm}	53.9%
1v_1p	F_{all_d}	54.5%	2v_4p	F_{all}	54.9%	4v_3p	F_{mm_d}	53.0%
1v_1p	F_{all_r}	55.1%	2v_5p	F_{mm}	54.3%	4v_3p	F_{mm_r}	50.3%
1v_1p	F_{all}	58.1%	2v_5p	F_{mm_d}	51.8%	4v_3p	$F_{d\_q2vp\_q3n}$	51.1%
1v_2p	F_{mm}	55.9%	2v_5p	F_{mm_r}	54.0%	4v_3p	$F_{r\_q2vp\_q3n}$	53.9%
1v_2p	F_{mm_d}	51.4%	2v_5p	$F_{d\_q2vp\_q3n}$	47.7%	4v_3p	F_{all_d}	53.9%
1v_2p	F_{mm_r}	53.8%	2v_5p	$F_{r\_q2vp\_q3n}$	54.0%	4v_3p	F_{all_r}	49.3%
1v_2p	$F_{d\_q2vp\_q3n}$	52.2%	2v_5p	F_{all_d}	54.5%	4v_3p	F_{all}	53.0%
1v_2p	$F_{r\_q2vp\_q3n}$	55.5%	2v_5p	F_{all_r}	52.8%	4v_4p	F_{mm}	53.9%

Continua na próxima página

Tabela A.3 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_2p	F_{all_d}	53.2%	2v_5p	F_{all}	53.5%	4v_4p	F_{mm_d}	53.0%
1v_2p	F_{all_r}	54.2%	3v_1p	F_{mm}	55.0%	4v_4p	F_{mm_r}	50.3%
1v_2p	F_{all}	56.9%	3v_1p	F_{mm_d}	55.1%	4v_4p	$F_{d\_q2vp\_q3n}$	51.1%
1v_3p	F_{mm}	53.7%	3v_1p	F_{mm_r}	55.0%	4v_4p	$F_{r\_q2vp\_q3n}$	53.9%
1v_3p	F_{mm_d}	56.5%	3v_1p	$F_{d\_q2vp\_q3n}$	51.7%	4v_4p	F_{all_d}	53.9%
1v_3p	F_{mm_r}	53.5%	3v_1p	$F_{r\_q2vp\_q3n}$	52.2%	4v_4p	F_{all_r}	49.3%
1v_3p	$F_{d\_q2vp\_q3n}$	50.5%	3v_1p	F_{all_d}	55.6%	4v_4p	F_{all}	53.0%
1v_3p	$F_{r\_q2vp\_q3n}$	56.9%	3v_1p	F_{all_r}	52.5%	4v_5p	F_{mm}	57.4%
1v_3p	F_{all_d}	54.1%	3v_1p	F_{all}	53.6%	4v_5p	F_{mm_d}	52.4%
1v_3p	F_{all_r}	53.5%	3v_2p	F_{mm}	55.0%	4v_5p	F_{mm_r}	51.4%
1v_3p	F_{all}	55.0%	3v_2p	F_{mm_d}	55.1%	4v_5p	$F_{d\_q2vp\_q3n}$	53.6%
1v_4p	F_{mm}	55.2%	3v_2p	F_{mm_r}	55.0%	4v_5p	$F_{r\_q2vp\_q3n}$	55.1%
1v_4p	F_{mm_d}	55.8%	3v_2p	$F_{d\_q2vp\_q3n}$	51.7%	4v_5p	F_{all_d}	54.7%
1v_4p	F_{mm_r}	54.8%	3v_2p	$F_{r\_q2vp\_q3n}$	52.2%	4v_5p	F_{all_r}	48.6%
1v_4p	$F_{d\_q2vp\_q3n}$	49.5%	3v_2p	F_{all_d}	55.6%	4v_5p	F_{all}	56.0%
1v_4p	$F_{r\_q2vp\_q3n}$	53.9%	3v_2p	F_{all_r}	52.5%	5v_1p	F_{mm}	53.2%
1v_4p	F_{all_d}	54.3%	3v_2p	F_{all}	53.6%	5v_1p	F_{mm_d}	51.0%
1v_4p	F_{all_r}	54.5%	3v_3p	F_{mm}	55.0%	5v_1p	F_{mm_r}	54.2%
1v_4p	F_{all}	55.8%	3v_3p	F_{mm_d}	55.1%	5v_1p	$F_{d\_q2vp\_q3n}$	43.6%
1v_5p	F_{mm}	53.9%	3v_3p	F_{mm_r}	55.0%	5v_1p	$F_{r\_q2vp\_q3n}$	47.9%
1v_5p	F_{mm_d}	57.2%	3v_3p	$F_{d\_q2vp\_q3n}$	51.7%	5v_1p	F_{all_d}	49.0%
1v_5p	F_{mm_r}	52.8%	3v_3p	$F_{r\_q2vp\_q3n}$	52.2%	5v_1p	F_{all_r}	53.2%
1v_5p	$F_{d\_q2vp\_q3n}$	50.0%	3v_3p	F_{all_d}	55.6%	5v_1p	F_{all}	53.2%
1v_5p	$F_{r\_q2vp\_q3n}$	54.8%	3v_3p	F_{all_r}	52.5%	5v_2p	F_{mm}	53.2%
1v_5p	F_{all_d}	54.6%	3v_3p	F_{all}	53.6%	5v_2p	F_{mm_d}	51.0%
1v_5p	F_{all_r}	52.4%	3v_4p	F_{mm}	55.1%	5v_2p	F_{mm_r}	54.2%
1v_5p	F_{all}	54.6%	3v_4p	F_{mm_d}	55.3%	5v_2p	$F_{d\_q2vp\_q3n}$	43.6%
2v_1p	F_{mm}	54.9%	3v_4p	F_{mm_r}	53.6%	5v_2p	$F_{r\_q2vp\_q3n}$	47.9%
2v_1p	F_{mm_d}	55.0%	3v_4p	$F_{d\_q2vp\_q3n}$	50.0%	5v_2p	F_{all_d}	49.0%
2v_1p	F_{mm_r}	54.0%	3v_4p	$F_{r\_q2vp\_q3n}$	53.7%	5v_2p	F_{all_r}	53.2%
2v_1p	$F_{d\_q2vp\_q3n}$	48.2%	3v_4p	F_{all_d}	58.4%	5v_2p	F_{all}	53.2%
2v_1p	$F_{r\_q2vp\_q3n}$	54.2%	3v_4p	F_{all_r}	51.7%	5v_3p	F_{mm}	53.2%
2v_1p	F_{all_d}	54.7%	3v_4p	F_{all}	55.0%	5v_3p	F_{mm_d}	51.0%
2v_1p	F_{all_r}	52.8%	3v_5p	F_{mm}	55.7%	5v_3p	F_{mm_r}	54.2%
2v_1p	F_{all}	53.3%	3v_5p	F_{mm_d}	54.2%	5v_3p	$F_{d\_q2vp\_q3n}$	43.6%
2v_2p	F_{mm}	54.4%	3v_5p	F_{mm_r}	52.8%	5v_3p	$F_{r\_q2vp\_q3n}$	47.9%
2v_2p	F_{mm_d}	56.1%	3v_5p	$F_{d\_q2vp\_q3n}$	48.3%	5v_3p	F_{all_d}	49.0%
2v_2p	F_{mm_r}	54.6%	3v_5p	$F_{r\_q2vp\_q3n}$	50.9%	5v_3p	F_{all_r}	53.2%
2v_2p	$F_{d\_q2vp\_q3n}$	49.2%	3v_5p	F_{all_d}	54.5%	5v_3p	F_{all}	53.2%
2v_2p	$F_{r\_q2vp\_q3n}$	54.8%	3v_5p	F_{all_r}	48.6%	5v_4p	F_{mm}	53.2%
2v_2p	F_{all_d}	56.5%	3v_5p	F_{all}	53.9%	5v_4p	F_{mm_d}	51.0%
2v_2p	F_{all_r}	52.7%	4v_1p	F_{mm}	53.9%	5v_4p	F_{mm_r}	54.2%
2v_2p	F_{all}	55.1%	4v_1p	F_{mm_d}	53.0%	5v_4p	$F_{d\_q2vp\_q3n}$	43.6%

Continua na próxima página

Tabela A.3 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
2v_3p	F_{mm}	54.9%	4v_1p	F_{mm_r}	50.3%	5v_4p	$F_{r\_q2vp\_q3n}$	47.9%
2v_3p	F_{mm_d}	53.0%	4v_1p	$F_{d\_q2vp\_q3n}$	51.1%	5v_4p	F_{all_d}	49.0%
2v_3p	F_{mm_r}	53.0%	4v_1p	$F_{r\_q2vp\_q3n}$	53.9%	5v_4p	F_{all_r}	53.2%
2v_3p	$F_{d\_q2vp\_q3n}$	48.0%	4v_1p	F_{all_d}	53.9%	5v_4p	F_{all}	53.2%
2v_3p	$F_{r\_q2vp\_q3n}$	55.3%	4v_1p	F_{all_r}	49.3%	5v_5p	F_{mm}	53.2%
2v_3p	F_{all_d}	54.9%	4v_1p	F_{all}	53.0%	5v_5p	F_{mm_d}	51.0%
2v_3p	F_{all_r}	53.0%	4v_2p	F_{mm}	53.9%	5v_5p	F_{mm_r}	54.2%
2v_3p	F_{all}	53.7%	4v_2p	F_{mm_d}	53.0%	5v_5p	$F_{d\_q2vp\_q3n}$	43.6%
2v_4p	F_{mm}	53.5%	4v_2p	F_{mm_r}	50.3%	5v_5p	$F_{r\_q2vp\_q3n}$	47.9%
2v_4p	F_{mm_d}	53.9%	4v_2p	$F_{d\_q2vp\_q3n}$	51.1%	5v_5p	F_{all_d}	49.0%
						5v_5p	F_{all_r}	53.2%
						5v_5p	F_{all}	53.2%

Tabela A.4: Acurácia média (ACC) resultante dos experimentos para o *XGBoost*, para cada combinação (Comb.) e *feature*

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_1p	F_{mm}	54.5%	2v_4p	F_{mm_r}	49.8%	4v_2p	$F_{r\_q2vp\_q3n}$	56.7%
1v_1p	F_{mm_d}	50.5%	2v_4p	$F_{d\_q2vp\_q3n}$	48.9%	4v_2p	F_{all_d}	52.1%
1v_1p	F_{mm_r}	52.3%	2v_4p	$F_{r\_q2vp\_q3n}$	53.5%	4v_2p	F_{all_r}	53.9%
1v_1p	$F_{d\_q2vp\_q3n}$	48.7%	2v_4p	F_{all_d}	49.1%	4v_2p	F_{all}	50.7%
1v_1p	$F_{r\_q2vp\_q3n}$	53.4%	2v_4p	F_{all_r}	52.1%	4v_3p	F_{mm}	55.3%
1v_1p	F_{all_d}	50.7%	2v_4p	F_{all}	51.8%	4v_3p	F_{mm_d}	53.9%
1v_1p	F_{all_r}	54.3%	2v_5p	F_{mm}	55.7%	4v_3p	F_{mm_r}	44.8%
1v_1p	F_{all}	52.3%	2v_5p	F_{mm_d}	56.0%	4v_3p	$F_{d\_q2vp\_q3n}$	47.0%
1v_2p	F_{mm}	56.3%	2v_5p	F_{mm_r}	48.7%	4v_3p	$F_{r\_q2vp\_q3n}$	56.7%
1v_2p	F_{mm_d}	52.8%	2v_5p	$F_{d\_q2vp\_q3n}$	48.4%	4v_3p	F_{all_d}	52.1%
1v_2p	F_{mm_r}	51.8%	2v_5p	$F_{r\_q2vp\_q3n}$	51.5%	4v_3p	F_{all_r}	53.9%
1v_2p	$F_{d\_q2vp\_q3n}$	49.5%	2v_5p	F_{all_d}	50.1%	4v_3p	F_{all}	50.7%
1v_2p	$F_{r\_q2vp\_q3n}$	54.3%	2v_5p	F_{all_r}	47.7%	4v_4p	F_{mm}	55.3%
1v_2p	F_{all_d}	47.9%	2v_5p	F_{all}	53.8%	4v_4p	F_{mm_d}	53.9%
1v_2p	F_{all_r}	46.6%	3v_1p	F_{mm}	54.5%	4v_4p	F_{mm_r}	44.8%
1v_2p	F_{all}	53.2%	3v_1p	F_{mm_d}	52.0%	4v_4p	$F_{d\_q2vp\_q3n}$	47.0%
1v_3p	F_{mm}	51.4%	3v_1p	F_{mm_r}	51.4%	4v_4p	$F_{r\_q2vp\_q3n}$	56.7%
1v_3p	F_{mm_d}	49.3%	3v_1p	$F_{d\_q2vp\_q3n}$	49.5%	4v_4p	F_{all_d}	52.1%
1v_3p	F_{mm_r}	51.8%	3v_1p	$F_{r\_q2vp\_q3n}$	50.8%	4v_4p	F_{all_r}	53.9%
1v_3p	$F_{d\_q2vp\_q3n}$	47.8%	3v_1p	F_{all_d}	53.9%	4v_4p	F_{all}	50.7%
1v_3p	$F_{r\_q2vp\_q3n}$	54.2%	3v_1p	F_{all_r}	47.7%	4v_5p	F_{mm}	55.6%
1v_3p	F_{all_d}	49.1%	3v_1p	F_{all}	50.0%	4v_5p	F_{mm_d}	50.5%
1v_3p	F_{all_r}	48.8%	3v_2p	F_{mm}	54.5%	4v_5p	F_{mm_r}	50.5%
1v_3p	F_{all}	52.0%	3v_2p	F_{mm_d}	52.0%	4v_5p	$F_{d\_q2vp\_q3n}$	51.3%

Continua na próxima página

Tabela A.4 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_4p	F_{mm}	52.4%	3v_2p	F_{mm_r}	51.4%	4v_5p	$F_{r\_q2vp\_q3n}$	58.3%
1v_4p	F_{mm_d}	50.7%	3v_2p	$F_{d\_q2vp\_q3n}$	49.5%	4v_5p	F_{all_d}	54.2%
1v_4p	F_{mm_r}	51.7%	3v_2p	$F_{r\_q2vp\_q3n}$	50.8%	4v_5p	F_{all_r}	52.7%
1v_4p	$F_{d\_q2vp\_q3n}$	47.0%	3v_2p	F_{all_d}	53.9%	4v_5p	F_{all}	53.7%
1v_4p	$F_{r\_q2vp\_q3n}$	55.0%	3v_2p	F_{all_r}	47.7%	5v_1p	F_{mm}	50.0%
1v_4p	F_{all_d}	52.0%	3v_2p	F_{all}	50.0%	5v_1p	F_{mm_d}	46.7%
1v_4p	F_{all_r}	49.4%	3v_3p	F_{mm}	54.5%	5v_1p	F_{mm_r}	46.8%
1v_4p	F_{all}	51.1%	3v_3p	F_{mm_d}	52.0%	5v_1p	$F_{d\_q2vp\_q3n}$	46.8%
1v_5p	F_{mm}	55.9%	3v_3p	F_{mm_r}	51.4%	5v_1p	$F_{r\_q2vp\_q3n}$	56.4%
1v_5p	F_{mm_d}	53.5%	3v_3p	$F_{d\_q2vp\_q3n}$	49.5%	5v_1p	F_{all_d}	46.8%
1v_5p	F_{mm_r}	53.5%	3v_3p	$F_{r\_q2vp\_q3n}$	50.8%	5v_1p	F_{all_r}	52.1%
1v_5p	$F_{d\_q2vp\_q3n}$	47.4%	3v_3p	F_{all_d}	53.9%	5v_1p	F_{all}	45.6%
1v_5p	$F_{r\_q2vp\_q3n}$	53.5%	3v_3p	F_{all_r}	47.7%	5v_2p	F_{mm}	50.0%
1v_5p	F_{all_d}	50.4%	3v_3p	F_{all}	50.0%	5v_2p	F_{mm_d}	46.7%
1v_5p	F_{all_r}	49.8%	3v_4p	F_{mm}	52.0%	5v_2p	F_{mm_r}	46.8%
1v_5p	F_{all}	51.1%	3v_4p	F_{mm_d}	54.2%	5v_2p	$F_{d\_q2vp\_q3n}$	46.8%
2v_1p	F_{mm}	55.7%	3v_4p	F_{mm_r}	52.0%	5v_2p	$F_{r\_q2vp\_q3n}$	56.4%
2v_1p	F_{mm_d}	50.2%	3v_4p	$F_{d\_q2vp\_q3n}$	47.2%	5v_2p	F_{all_d}	46.8%
2v_1p	F_{mm_r}	50.2%	3v_4p	$F_{r\_q2vp\_q3n}$	54.8%	5v_2p	F_{all_r}	52.1%
2v_1p	$F_{d\_q2vp\_q3n}$	45.5%	3v_4p	F_{all_d}	52.5%	5v_2p	F_{all}	45.6%
2v_1p	$F_{r\_q2vp\_q3n}$	54.5%	3v_4p	F_{all_r}	51.2%	5v_3p	F_{mm}	50.0%
2v_1p	F_{all_d}	51.2%	3v_4p	F_{all}	51.1%	5v_3p	F_{mm_d}	46.7%
2v_1p	F_{all_r}	51.7%	3v_5p	F_{mm}	54.2%	5v_3p	F_{mm_r}	46.8%
2v_1p	F_{all}	51.6%	3v_5p	F_{mm_d}	55.7%	5v_3p	$F_{d\_q2vp\_q3n}$	46.8%
2v_2p	F_{mm}	57.7%	3v_5p	F_{mm_r}	50.6%	5v_3p	$F_{r\_q2vp\_q3n}$	56.4%
2v_2p	F_{mm_d}	51.5%	3v_5p	$F_{d\_q2vp\_q3n}$	48.9%	5v_3p	F_{all_d}	46.8%
2v_2p	F_{mm_r}	52.0%	3v_5p	$F_{r\_q2vp\_q3n}$	52.0%	5v_3p	F_{all_r}	52.1%
2v_2p	$F_{d\_q2vp\_q3n}$	46.4%	3v_5p	F_{all_d}	54.3%	5v_3p	F_{all}	45.6%
2v_2p	$F_{r\_q2vp\_q3n}$	56.7%	3v_5p	F_{all_r}	48.3%	5v_4p	F_{mm}	50.0%
2v_2p	F_{all_d}	49.7%	3v_5p	F_{all}	52.3%	5v_4p	F_{mm_d}	46.7%
2v_2p	F_{all_r}	52.5%	4v_1p	F_{mm}	55.3%	5v_4p	F_{mm_r}	46.8%
2v_2p	F_{all}	54.1%	4v_1p	F_{mm_d}	53.9%	5v_4p	$F_{d\_q2vp\_q3n}$	46.8%
2v_3p	F_{mm}	55.6%	4v_1p	F_{mm_r}	44.8%	5v_4p	$F_{r\_q2vp\_q3n}$	56.4%
2v_3p	F_{mm_d}	52.3%	4v_1p	$F_{d\_q2vp\_q3n}$	47.0%	5v_4p	F_{all_d}	46.8%
2v_3p	F_{mm_r}	50.8%	4v_1p	$F_{r\_q2vp\_q3n}$	56.7%	5v_4p	F_{all_r}	52.1%
2v_3p	$F_{d\_q2vp\_q3n}$	48.7%	4v_1p	F_{all_d}	52.1%	5v_4p	F_{all}	45.6%
2v_3p	$F_{r\_q2vp\_q3n}$	55.8%	4v_1p	F_{all_r}	53.9%	5v_5p	F_{mm}	50.0%
2v_3p	F_{all_d}	51.8%	4v_1p	F_{all}	50.7%	5v_5p	F_{mm_d}	46.7%
2v_3p	F_{all_r}	52.3%	4v_2p	F_{mm}	55.3%	5v_5p	F_{mm_r}	46.8%
2v_3p	F_{all}	52.5%	4v_2p	F_{mm_d}	53.9%	5v_5p	$F_{d\_q2vp\_q3n}$	46.8%
2v_4p	F_{mm}	56.1%	4v_2p	F_{mm_r}	44.8%	5v_5p	$F_{r\_q2vp\_q3n}$	56.4%
2v_4p	F_{mm_d}	54.4%	4v_2p	$F_{d\_q2vp\_q3n}$	47.0%	5v_5p	F_{all_d}	46.8%
						5v_5p	F_{all_r}	52.1%

Continua na próxima página

Tabela A.4 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
						5v_5p	F_{all}	45.6%

Tabela A.5: Acurácia média (ACC) resultante dos experimentos para o *MLP Classifier*, para cada combinação (Comb.) e *feature*

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_1p	F_{mm}	50.7%	2v_4p	F_{mm_r}	53.7%	4v_2p	$F_{r\_q2vp\_q3n}$	54.4%
1v_1p	F_{mm_d}	51.1%	2v_4p	$F_{d\_q2vp\_q3n}$	53.0%	4v_2p	F_{all_d}	48.0%
1v_1p	F_{mm_r}	53.9%	2v_4p	$F_{r\_q2vp\_q3n}$	54.7%	4v_2p	F_{all_r}	50.3%
1v_1p	$F_{d\_q2vp\_q3n}$	51.5%	2v_4p	F_{all_d}	50.9%	4v_2p	F_{all}	52.6%
1v_1p	$F_{r\_q2vp\_q3n}$	52.7%	2v_4p	F_{all_r}	49.6%	4v_3p	F_{mm}	48.4%
1v_1p	F_{all_d}	50.3%	2v_4p	F_{all}	49.2%	4v_3p	F_{mm_d}	53.0%
1v_1p	F_{all_r}	51.9%	2v_5p	F_{mm}	51.3%	4v_3p	F_{mm_r}	47.6%
1v_1p	F_{all}	53.7%	2v_5p	F_{mm_d}	50.8%	4v_3p	$F_{d\_q2vp\_q3n}$	53.4%
1v_2p	F_{mm}	52.6%	2v_5p	F_{mm_r}	53.0%	4v_3p	$F_{r\_q2vp\_q3n}$	54.4%
1v_2p	F_{mm_d}	49.5%	2v_5p	$F_{d\_q2vp\_q3n}$	54.2%	4v_3p	F_{all_d}	48.0%
1v_2p	F_{mm_r}	53.0%	2v_5p	$F_{r\_q2vp\_q3n}$	53.0%	4v_3p	F_{all_r}	50.3%
1v_2p	$F_{d\_q2vp\_q3n}$	51.8%	2v_5p	F_{all_d}	49.4%	4v_3p	F_{all}	52.6%
1v_2p	$F_{r\_q2vp\_q3n}$	54.0%	2v_5p	F_{all_r}	49.6%	4v_4p	F_{mm}	48.4%
1v_2p	F_{all_d}	51.2%	2v_5p	F_{all}	51.6%	4v_4p	F_{mm_d}	53.0%
1v_2p	F_{all_r}	52.4%	3v_1p	F_{mm}	48.9%	4v_4p	F_{mm_r}	47.6%
1v_2p	F_{all}	46.4%	3v_1p	F_{mm_d}	50.0%	4v_4p	$F_{d\_q2vp\_q3n}$	53.4%
1v_3p	F_{mm}	51.4%	3v_1p	F_{mm_r}	50.0%	4v_4p	$F_{r\_q2vp\_q3n}$	54.4%
1v_3p	F_{mm_d}	49.2%	3v_1p	$F_{d\_q2vp\_q3n}$	50.8%	4v_4p	F_{all_d}	48.0%
1v_3p	F_{mm_r}	53.7%	3v_1p	$F_{r\_q2vp\_q3n}$	52.5%	4v_4p	F_{all_r}	50.3%
1v_3p	$F_{d\_q2vp\_q3n}$	53.7%	3v_1p	F_{all_d}	51.7%	4v_4p	F_{all}	52.6%
1v_3p	$F_{r\_q2vp\_q3n}$	53.5%	3v_1p	F_{all_r}	47.2%	4v_5p	F_{mm}	50.5%
1v_3p	F_{all_d}	51.4%	3v_1p	F_{all}	52.2%	4v_5p	F_{mm_d}	53.2%
1v_3p	F_{all_r}	49.5%	3v_2p	F_{mm}	48.9%	4v_5p	F_{mm_r}	45.4%
1v_3p	F_{all}	49.7%	3v_2p	F_{mm_d}	50.0%	4v_5p	$F_{d\_q2vp\_q3n}$	53.6%
1v_4p	F_{mm}	52.6%	3v_2p	F_{mm_r}	50.0%	4v_5p	$F_{r\_q2vp\_q3n}$	54.2%
1v_4p	F_{mm_d}	49.4%	3v_2p	$F_{d\_q2vp\_q3n}$	50.8%	4v_5p	F_{all_d}	49.5%
1v_4p	F_{mm_r}	53.2%	3v_2p	$F_{r\_q2vp\_q3n}$	52.5%	4v_5p	F_{all_r}	50.5%
1v_4p	$F_{d\_q2vp\_q3n}$	52.8%	3v_2p	F_{all_d}	51.7%	4v_5p	F_{all}	54.5%
1v_4p	$F_{r\_q2vp\_q3n}$	52.5%	3v_2p	F_{all_r}	47.2%	5v_1p	F_{mm}	54.2%
1v_4p	F_{all_d}	53.2%	3v_2p	F_{all}	52.2%	5v_1p	F_{mm_d}	48.9%
1v_4p	F_{all_r}	50.0%	3v_3p	F_{mm}	48.9%	5v_1p	F_{mm_r}	55.3%
1v_4p	F_{all}	52.8%	3v_3p	F_{mm_d}	50.0%	5v_1p	$F_{d\_q2vp\_q3n}$	54.2%
1v_5p	F_{mm}	50.2%	3v_3p	F_{mm_r}	50.0%	5v_1p	$F_{r\_q2vp\_q3n}$	41.5%
1v_5p	F_{mm_d}	52.8%	3v_3p	$F_{d\_q2vp\_q3n}$	50.8%	5v_1p	F_{all_d}	49.0%
1v_5p	F_{mm_r}	50.0%	3v_3p	$F_{r\_q2vp\_q3n}$	52.5%	5v_1p	F_{all_r}	39.4%

Continua na próxima página

Tabela A.5 – continuação da página anterior

Comb.	*feature*	ACC	Comb.	*feature*	ACC	Comb.	*feature*	ACC
1v_5p	$F_{d\_q2vp\_q3n}$	52.6%	3v_3p	F_{all_d}	51.7%	5v_1p	F_{all}	43.6%
1v_5p	$F_{r\_q2vp\_q3n}$	54.6%	3v_3p	F_{all_r}	47.2%	5v_2p	F_{mm}	54.2%
1v_5p	F_{all_d}	51.5%	3v_3p	F_{all}	52.2%	5v_2p	F_{mm_d}	48.9%
1v_5p	F_{all_r}	53.9%	3v_4p	F_{mm}	45.8%	5v_2p	F_{mm_r}	55.3%
1v_5p	F_{all}	50.4%	3v_4p	F_{mm_d}	50.0%	5v_2p	$F_{d\_q2vp\_q3n}$	54.2%
2v_1p	F_{mm}	51.9%	3v_4p	F_{mm_r}	47.7%	5v_2p	$F_{r\_q2vp\_q3n}$	41.5%
2v_1p	F_{mm_d}	54.5%	3v_4p	$F_{d\_q2vp\_q3n}$	53.7%	5v_2p	F_{all_d}	49.0%
2v_1p	F_{mm_r}	53.3%	3v_4p	$F_{r\_q2vp\_q3n}$	53.1%	5v_2p	F_{all_r}	39.4%
2v_1p	$F_{d\_q2vp\_q3n}$	53.6%	3v_4p	F_{all_d}	51.7%	5v_2p	F_{all}	43.6%
2v_1p	$F_{r\_q2vp\_q3n}$	51.7%	3v_4p	F_{all_r}	47.5%	5v_3p	F_{mm}	54.2%
2v_1p	F_{all_d}	51.5%	3v_4p	F_{all}	47.7%	5v_3p	F_{mm_d}	48.9%
2v_1p	F_{all_r}	45.8%	3v_5p	F_{mm}	49.2%	5v_3p	F_{mm_r}	55.3%
2v_1p	F_{all}	49.1%	3v_5p	F_{mm_d}	52.9%	5v_3p	$F_{d\_q2vp\_q3n}$	54.2%
2v_2p	F_{mm}	50.4%	3v_5p	F_{mm_r}	48.6%	5v_3p	$F_{r\_q2vp\_q3n}$	41.5%
2v_2p	F_{mm_d}	53.2%	3v_5p	$F_{d\_q2vp\_q3n}$	52.9%	5v_3p	F_{all_d}	49.0%
2v_2p	F_{mm_r}	52.0%	3v_5p	$F_{r\_q2vp\_q3n}$	54.6%	5v_3p	F_{all_r}	39.4%
2v_2p	$F_{d\_q2vp\_q3n}$	52.5%	3v_5p	F_{all_d}	53.7%	5v_3p	F_{all}	43.6%
2v_2p	$F_{r\_q2vp\_q3n}$	52.3%	3v_5p	F_{all_r}	48.6%	5v_4p	F_{mm}	54.2%
2v_2p	F_{all_d}	53.0%	3v_5p	F_{all}	51.7%	5v_4p	F_{mm_d}	48.9%
2v_2p	F_{all_r}	47.6%	4v_1p	F_{mm}	48.4%	5v_4p	F_{mm_r}	55.3%
2v_2p	F_{all}	49.2%	4v_1p	F_{mm_d}	53.0%	5v_4p	$F_{d\_q2vp\_q3n}$	54.2%
2v_3p	F_{mm}	51.1%	4v_1p	F_{mm_r}	47.6%	5v_4p	$F_{r\_q2vp\_q3n}$	41.5%
2v_3p	F_{mm_d}	49.0%	4v_1p	$F_{d\_q2vp\_q3n}$	53.4%	5v_4p	F_{all_d}	49.0%
2v_3p	F_{mm_r}	50.3%	4v_1p	$F_{r\_q2vp\_q3n}$	54.4%	5v_4p	F_{all_r}	39.4%
2v_3p	$F_{d\_q2vp\_q3n}$	52.5%	4v_1p	F_{all_d}	48.0%	5v_4p	F_{all}	43.6%
2v_3p	$F_{r\_q2vp\_q3n}$	50.1%	4v_1p	F_{all_r}	50.3%	5v_5p	F_{mm}	54.2%
2v_3p	F_{all_d}	48.7%	4v_1p	F_{all}	52.6%	5v_5p	F_{mm_d}	48.9%
2v_3p	F_{all_r}	45.7%	4v_2p	F_{mm}	48.4%	5v_5p	F_{mm_r}	55.3%
2v_3p	F_{all}	48.7%	4v_2p	F_{mm_d}	53.0%	5v_5p	$F_{d\_q2vp\_q3n}$	54.2%
2v_4p	F_{mm}	50.3%	4v_2p	F_{mm_r}	47.6%	5v_5p	$F_{r\_q2vp\_q3n}$	41.5%
2v_4p	F_{mm_d}	53.0%	4v_2p	$F_{d\_q2vp\_q3n}$	53.4%	5v_5p	F_{all_d}	49.0%
						5v_5p	F_{all_r}	39.4%
						5v_5p	F_{all}	43.6%