

# Classificador de Qualidade de Voz em Sistemas de Comunicações baseado em Deep Learning

Lucas Hilário da Costa, Andreza Patrícia Batista, Renata Lopes Rosa e Demóstenes Zegarra Rodríguez

**Resumo**— Neste trabalho, dois modelos de deep learning (DL) são implementados para classificar a qualidade do sinal de voz transmitido em uma comunicação VoIP afetada pela perda de pacotes. O primeiro modelo estima a taxa de perda de pacotes (PLR) de sinais degradados e o segundo modelo estima a qualidade dos sinais de voz, classificando-os em quatro classes. Assim, foram elaborados dois bancos de dados, cada um contendo quatro classes distintas, utilizando arquivos de recomendação ITU-T P.501 de acordo com o índice Mean Opinion Score (MOS) de cada arquivo degradado. Observou-se que o modelo que estima o PLR teve 96% de acurácia, enquanto o modelo que estima o índice MOS atingiu 97% de acurácia. É importante destacar que o algoritmo ITU-T P.563 atingiu uma acurácia de 61,83%, a qual é inferior ao alcançado pelo modelo de estimação de MOS proposto, sendo que ambos modelos são não intrusivos.

**Palavras-Chave**— Qualidade de Voz, ITU-T P.862, ITU-T P.563, Deep Learning, Machine Learning.

**Abstract**— In this work, in order to classify the quality of the voice signal transmitted in a VoIP communication affected by packet loss, two deep learning (DL) models were implemented. The first model estimates the packet loss rate (PLR) of degraded signals, and the second model estimates the quality of voice signals, classifying them into four classes. Thus, two databases were prepared, each containing four distinct classes, using ITU-T P.501 recommendation files according to the Mean Opinion Score (MOS) index of each degraded file. It was observed that the model that estimates the PLR was 96% accurate, while the model that estimates the MOS reached 97% of accuracy. It is important to highlight that the ITU-T P.563 algorithm reached an accuracy of 61.83%, which is lower than that achieved by the proposed MOS estimation model, both models being non-intrusive.

**Palavras-Chave**— Voice quality, ITU-T P.862, ITU-T P.563, ITU-T P.501, Deep Learning, Machine Learning.

## I. INTRODUÇÃO

Perda de pacotes em um sistema de comunicações afetam diretamente a qualidade do sinal transmitido. Essas perdas, podem estar ligadas ao alto tráfego de dados, roteadores congestionados e/ou problemas nos meios físicos de transporte.

Para avaliar a qualidade do sinal de voz causado por fatores de degradação da rede, existem métodos bem definidos, que são especificados através das recomendações técnicas da *International Telecommunication Union - Telecommunication Standardization Sector* (ITU-T). Essas recomendações definem como será avaliada a qualidade de voz, pontuando a

Lucas Hilario da Costa, Renata Lopes Rosa e Demóstenes Zegarra Rodríguez, Departamento de Ciência da Computação, Universidade Federal de Lavras, Lavras-MG, e-mails: lucash@gmail.com, renta.rosa@ufla.br, demostenes.zegarra@ufla.br. Andreza Patrícia Batista, Programa de Pós-Graduação em Engenharia de Sistema e Automação, Universidade Federal de Lavras, Lavras-MG, e Setor de Laboratórios, Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais-Campus Formiga, Formiga-MG,

qualidade das comunicações e são de grande importância para o monitoramento da Qualidade da Experiência (QoE) dos usuários do serviço.

Os métodos de avaliação da qualidade da voz podem ser classificados em métodos subjetivos ou objetivos. No método subjetivo, várias pessoas são solicitadas a avaliar a qualidade das amostras de voz, no qual resulta na Índice MOS [1], nota atribuída à qualidade de voz pela experiência do usuário. No entanto, é um método demorado e caro para ser usado no monitoramento contínuo da QoE do sistema em um cenário de comunicação de voz real. Por outro lado, os métodos objetivos utilizam algoritmos e tentam prever de forma aproximada o escore de QoE do usuário, que seria dado em testes subjetivos por indivíduos. Além disso, os métodos objetivos são subdivididos em dois métodos, intrusivos, e não intrusivos, cujos algoritmos mais representativos estão descritos nas ITU-T Recs. P.862 [2] e P.563 [3], respectivamente. Na atualidade existem outras propostas para avaliar a qualidade do sinal de voz, sendo que alguns deles utilizam algoritmos de aprendizado de máquina.

A rede neural convolucional (CNN) é uma classe de rede neural profunda que explora a forte correlação local e espacial em imagens naturais, alcançando grande desempenho no campo da análise visual. Além disso, as CNNs têm sido usadas no campo do processamento acústico e provaram ser capazes de aprender o padrão espectral-temporal do som e diferenciá-lo para fins de classificação [4]. Nos últimos anos, vários algoritmos de classificação de cenas acústicas foram propostos onde os algoritmos mais utilizados incluem Support-Vector Machine (SVM), Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), entre outros.

O principal objetivo deste trabalho é desenvolver um modelo não intrusivo para classificação de qualidade de voz em comunicação VoIP (Voz sobre Protocolo de Internet) utilizando algoritmos de Deep Learning. Assim, foram implementados dois modelos: (i) o primeiro capaz de classificar o grau de degradação da voz através do índice MOS, e considerando 4 classes de qualidade e (ii) o segundo capaz de determinar a taxa de perda de pacotes do sinal de voz.

O restante do artigo está organizado como segue: A Seção II apresenta alguns trabalhos relacionados; a metodologia e descrição das etapas do trabalho são descritas na Seção III. A Seção IV apresenta os resultados experimentais obtidos e discussões sobre eles. Finalmente, as conclusões do trabalho são apresentadas na Seção V.

## II. TRABALHOS RELACIONADOS

Em [5], autores utilizaram as CNN's para detectar a quantidade de pacotes perdidos em sistemas de comunicação de voz. Para estimar a quantidade de pacotes perdidos, utilizaram espectrogramas dos sinais de fala transmitidos como entrada de uma rede neural convolucional. O modelo proposto, permitiu estimar a taxa de perda de pacotes de um sistema de comunicação usando arquivo de fala gravado no receptor, sem a necessidade do sinal de fala de referência. Os resultados mostraram que o modelo pode reduzir o erro de previsão em mais de 75% quando comparado a um modelo baseado em recursos do MFCC.

No trabalho de [6], foram utilizados vários cenários de rede que consideram diferentes taxas de perda de pacotes (PLR), nos quais, os sinais com problemas são avaliados usando o algoritmo descrito na recomendação ITU-T P.862. Os resultados mostraram uma relação entre os parâmetros de desvanecimento e PLR e o índice global de qualidade da fala. O objetivo principal do trabalho foi propor um modelo não intrusivo de classificação da qualidade da fala baseado em uma Deep Belief Network (DBN) que considera as deficiências com fio e sem fio no sinal da fala. Os resultados demonstraram uma alta correlação entre o modelo proposto baseado no algoritmo DBN e o da recomendação ITU-T P.862.

Em [7], um Deep Machine Listening para Estimating Speech Quality (DESQ) foi proposta, que consegue prever a qualidade de fala percebida com base nas probabilidades posteriores do fonema obtidas de uma rede neural profunda. A degradação dos fonemas é quantificada com a medida de Gini baseada em entropia que é comparada com a distância temporal média (MTD - Mean Temporal Distance) proposta anteriormente. Como as longas pausas de fala e perdas de pacote podem ter um grande efeito na qualidade da fala, o objetivo do trabalho era verificar se uma detecção de atividade de voz (VAD - Voice Activity Detection) possuía um efeito benéfico ou prejudicial sobre o poder preditivo do modelo proposto. A avaliação é realizada correlacionando a saída do modelo e os valores de MOS dos ouvintes com audição normal que classificaram os sinais degradados por artefatos típicos de uma transmissão VoIP. Em [8] é apresentada uma solução similar.

Em [9] introduziu-se várias tecnologias de processamento de perda de pacotes usadas para chamadas VoIP para reduzir a distorção fonética causada pela perda de pacotes em uma transmissão VoIP ocasionando a perda da qualidade da voz na transmissão.

## III. MATERIAIS E MÉTODOS

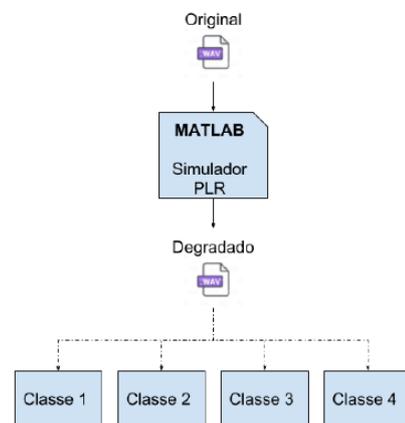
Para atingir os objetivos deste trabalho foram realizadas diferentes etapas: construção de base de dados a partir da degradação de áudios, determinação da topologia da rede.

O modelos de redes neurais foram desenvolvidos na linguagem Python utilizando-se o terminal do Ubuntu com o auxílio das APIs Keras e Tensorflow. Também, utilizou-se o LibROSA que é um pacote python para análise de sinais de áudio.

### A. Conjunto de dados

Para a criação das bases de dados, primeiro foi realizado downloads de bases de dados das recomendações ITU-T P.501 e ITU-T P.862, os dados foi preparados realizando um processamento no MATLAB que simula uma determinada taxa de perda de pacote, baseado no modelo de Gilbert-Elliot, no arquivo de voz original. Esse é uma das abordagens utilizadas para a determinação de perdas de pacotes em sistemas de transmissão [10] o qual utiliza uma cadeia de Markov de dois estados, denominados "Bom"(alta probabilidade de ocorrer e "Ruim"(baixa probabilidade de ocorrer). A Fig. 1 apresenta o procedimento realizado para obter os arquivos degradados.

Fig. 1: Geração de arquivos degradados para formação de conjunto de dados.



1) *Determinação das classes de PLR:* Para o modelo de classificação por perda de pacotes, foi realizado de acordo com Fig. 1. na qual são consideradas quatro classes distintas de acordo com sua taxa de PLR. As classes foram separadas da seguinte forma:

- Classe 1: PLR(0.5% + 1.0% + 1.5% + 2.0%)
- Classe 2: PLR(2.5% + 3.5% + 4.5% + 5.5%)
- Classe 3: PLR(7.0% + 8.0% + 9.0% + 10.0%)
- Classe 4: PLR(12.0% + 15.0% + 18.0% + 21.0%)

Cada classe possui um total de 1840 arquivos degradados.

Para gerar cada classe, foi executado o seguinte procedimento: para cada um dos 20 arquivos de áudio da base original foram realizadas o processamento com as 16 taxas de perda de pacote resultando num total de 320 arquivos para cada execução do algoritmo; para aumentar o número de arquivos e a diversificação entre eles o algoritmo foi executado 23 vezes. O resultado gerou um total de 7360 arquivos de áudio degradados, e como cada classe possui o mesmo número de arquivos degradados e é formada por 4 taxas de PLR diferentes, tem-se 1840 por classe.

2) *Determinação das classes de MOS:* Para o modelo de classificação por índice MOS foi executado, após o processamento dos arquivos, um novo algoritmo (também desenvolvido no MATLAB), capaz de automatizar a execução do algoritmo P.862 para todos os arquivos de áudio degradado, comparando-os com seus respectivos arquivos de áudio original. O resultado foi o índice MOS referente à perda de pacotes encontrada no arquivo de áudio degradado.

Após adquirir estes dados, os arquivos foram separados em 5 classes de acordo com seu índice MOS, utilizando como referência a tabela do Fator-R do E-Model representada na Tabela I.

TABELA I: Relação de valores R e índices MOS.

Fator R	Nível de Satisfação	MOS
90 a 1000	Muito satisfeito	4.3 - 4.5
80 a 90	Satisfeito	4.0 - 4.3
70 a 80	Alguns usuários insatisfeitos	3.6 - 4.0
60 a 70	Muitos usuários insatisfeitos	3.1 - 3.6
50 a 60	Praticamente todos insatisfeitos	2.8 - 3.1
0 a 50	Não recomendado	1.0 - 2.8

Porém, foram poucas as amostras pertencentes da classe 5 onde o índice MOS está entre 4,3 e 4,5, com isso foi adotado que os arquivos com índices maiores que 4,0 seriam somente uma classe. Também, arquivos com índices MOS menores a 3.1 foram agrupados na mesma classe. Assim, foram utilizadas as 4 seguintes classes:

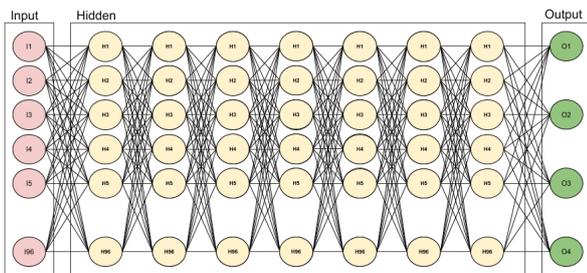
- Classe 1: MOS(1,00 a 3,09)
- Classe 2: MOS(3,10 a 3,59)
- Classe 3: MOS(3,60 a 3,99)
- Classe 4: MOS(4,00 a 4,50)

Cada classe possui um total de 2500 arquivos degradados.

**B. Topologia da Rede**

A arquitetura da Deep Learning utilizada no treinamento e validação do projeto foi definida de acordo com o número de características dos arquivos na entrada da rede neural, a qual é apresentada na Fig. 1:

Fig. 2: Arquitetura do modelo de Deep Learning.



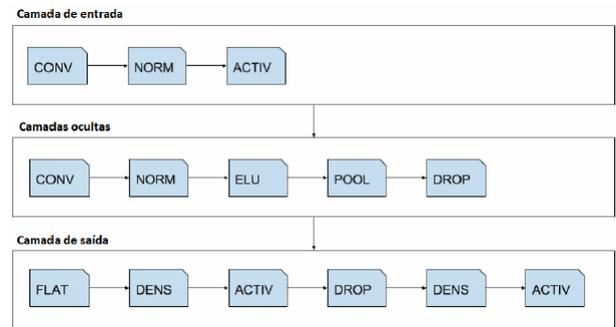
O modelo gerado possui as seguintes características:

- Possui uma camada de entrada (I); que é a responsável por realizar a entrada do arquivo de áudio na rede neural e com isso realizar uma leitura do número de características que a biblioteca Librosa pode analisar, sendo 96 o número de características que serão utilizadas na rede, definidas automaticamente pelo algoritmo.
- Possui sete camadas ocultas (H) com 96 neurônios cada, ligadas de forma recorrente de acordo com as RNN's. As camadas ocultas são responsáveis por analisar as características extraídas do arquivo de áudio para que seja possível realizar o reconhecimento e classificação das suas características.
- Possui uma camada de saída (O) com 4 neurônios, onde cada um representa uma saída ou classe. A camada de

saída é responsável por conectar todas as características e gerar o modelo final de classificação da rede.

A Fig. 3 apresenta a estrutura interna da rede neural proposta. As etapas representadas, são: Conv: é a realização de uma convolução; Norm: é a normalização dos dados; Activ: é a realização de uma função de ativação; Elu: é a realização de uma função de ativação Elu; Pool: é a realização da função Pooling para subdividir a imagem; Drop: é a chamada da função de Dropout, onde é desligado o percentual de neurônios da rede; Flat: é a realização da operação de Flatten, onde organiza os blocos separados pela operação de Pooling em um vetor linear; Dens: é a operação responsável por reduzir o número de neurônios agrupando os neurônios responsáveis por representar cada classe.

Fig. 3: Arquitetura do modelo de Deep Learning.

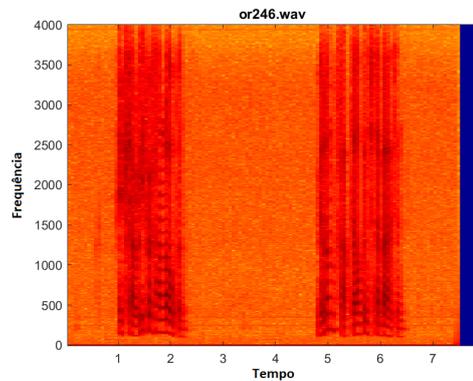


**IV. RESULTADOS E DISCUSSÕES**

Para realizar as etapas de treinamento e validação foi utilizado um computador com um processador I5 de 2.3 GHz, 16 GB de memória RAM e uma placa gráfica GTX 1050 TI com 4GB de memória de vídeo o tempo médio de execução do algoritmo de 9h.

Para a realização dos testes foram utilizados 80% dos arquivos de áudios para o treinamento e 20% para validação dos modelos. Para realizar a extração das características do arquivo de áudio, as CNN's geram principalmente os espectrogramas e os analisam como imagem pixel a pixel.

Fig. 4: Espectrograma do arquivo de áudio original.

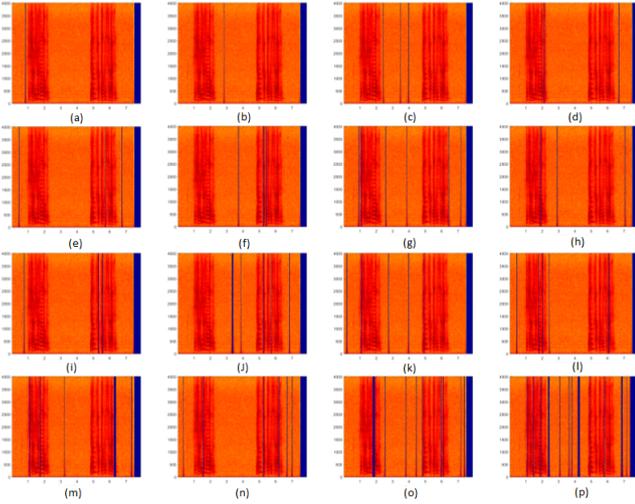


A Fig. 4 representa um espectrograma de um arquivo original, onde podemos verificar que o sinal da voz não sofreu

nenhum tipo de degradação, nesta a degradação é representada pela cor azul. Porém, a cor azul não representa somente a perda do sinal da voz, ela também representa o silêncio na conversação.

No entanto, o algoritmo da recomendação ITU-T P.862 possui a capacidade de diferenciar os silêncios e as perdas no sinal da voz.

Fig. 5: Espectrogramas dos arquivos de áudio degradados para os valores de PLR.



A Fig. 5 representa os espectrogramas gerados para o mesmo arquivo de áudio representado pela Fig. 4, porém como podemos ver a degradação ocorrida pelas taxas de perdas de pacotes definidas representaram uma degradação do sinal de voz para as diferentes taxas, e a medida que o percentual de perda de pacote aumenta podemos verificar que ocorre um aumento dos silêncios no sinal da voz. Na Fig. 5, as letras de "a" a "p" representam respectivamente os valores de PLR 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.5%, 4.5%, 5.5%, 7.0%, 8.0%, 9.0%, 10.0%, 12.0%, 15.0%, 18.0% e 21.0%.

#### A. Resultados utilizando o modelo para calcular o PLR

Os resultados obtidos nos testes com a base de dados preparada com os valores de PLR, foram conseguidos através da validação do modelo utilizando 20% do total de arquivos da base de dados, ou seja, 1472 arquivos sendo 368 arquivos para cada classe definida. Os resultados são mostrados em matrizes de confusão geradas para os números de épocas 100, 500 e 1000, e são apresentados nas Tabelas II, III e IV.

TABELA II: Resultado da matriz de confusão para 100 épocas para valores de PLR

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	356	12	0	0
Classe 2	10	319	39	0
Classe 3	0	20	321	27
Classe 4	0	0	49	319

A Tabela II apresenta os dados de acertos e erros dos testes realizados para 100 épocas de treinamento, através dela

podemos perceber que na validação do modelo treinado com 100 épocas tivemos um total de 157 erros de classificação de classes e um total de 1315 acertos na validação, obtendo uma acurácia de 89,334%.

TABELA III: Resultado da matriz de confusão para 500 épocas para valores de PLR

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	359	9	0	0
Classe 2	11	325	32	0
Classe 3	0	21	319	28
Classe 4	0	0	48	320

A Tabela III apresenta os dados de acertos e erros dos testes realizados para 500 épocas de treinamento, através dela podemos perceber que na validação do modelo treinado com 500 épocas tivemos um total de 149 erros de classificação de classes e um total de 1323 acertos na validação, obtendo uma acurácia de 89,877%.

TABELA IV: Resultado da matriz de confusão para 1000 épocas para valores de PLR

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	364	4	0	0
Classe 2	4	357	7	0
Classe 3	0	9	347	12
Classe 4	0	0	13	355

A Tabela IV apresenta os dados de acertos e erros dos testes realizados para 1000 épocas de treinamento, através dela podemos perceber que na validação do modelo treinado com 1000 épocas tivemos um total de 49 erros de classificação de classes e um total de 1423 acertos na validação, obtendo uma acurácia de 96,671%. Ao incrementar-se mais épocas não se obteve ganho significativo de desempenho.

#### B. Resultados utilizando o modelo MOS

Os resultados obtidos nos testes com a base de dados preparada com os valores do índice MOS foram obtidos através da validação do modelo utilizando 30% do total de arquivos da base de dados, onde cada classe contém um total de 750 arquivos para validação, resultando num total de 3000 arquivos para a validação do modelo.

Os resultados são mostrados em matrizes de confusão geradas para os seguintes números de épocas 100, 500 e 1000, e são apresentados nas Tabelas V, VI e VII.

TABELA V: Resultado da matriz de confusão para 100 épocas para valores de MOS

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	699	51	0	0
Classe 2	36	680	34	0
Classe 3	0	37	681	32
Classe 4	0	0	72	678

A Tabela V apresenta os dados de acertos e erros dos testes realizados para 100 épocas de treinamento, através dela

podemos perceber que na validação do modelo treinado com 100 épocas tivemos um total de 262 erros de classificação de classes e um total de 2736 acertos de classificação, obtendo uma acurácia de 91,200%.

TABELA VI: Resultado da matriz de confusão para 500 épocas para valores de MOS

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	716	34	0	0
Classe 2	31	694	25	0
Classe 3	0	28	705	17
Classe 4	0	0	47	703

A Tabela VI apresenta os dados de acertos e erros dos testes realizados para 500 épocas de treinamento, através dela podemos perceber que na validação do modelo treinado com 500 épocas tivemos um total de 182 erros de classificação de classes e um total de 2818 acertos de classificação, obtendo uma acurácia de 93,933%.

TABELA VII: Resultado da matriz de confusão para 1000 épocas para valores de MOS

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	735	15	0	0
Classe 2	10	729	11	0
Classe 3	0	13	728	9
Classe 4	0	0	17	733

A Tabela VII apresenta os dados de acertos e erros dos testes realizados para 1000 épocas de treinamento, através dela podemos perceber que na validação do modelo treinado com 1000 épocas tivemos um total de 75 erros de classificação de classes e um total de 2894 acertos de classificação, obtendo uma acurácia de 97,500%. Ao incrementar-se mais épocas não se obteve ganho significativo de desempenho.

### C. Comparação dos algoritmos P.563 vs P.862

Ao realizar a classificação dos índice MOS do Fator-R com os resultados dos algoritmos das recomendações ITU-T P.563 e ITU-T P.862, utilizando as 10,000 amostras, obtivemos, de acordo com a Tabela VIII, uma média de 61,83% de acerto quando comparamos os resultados do P.563 com os resultados do P.862. O algoritmo P.563 classificou de forma correta as classes de 6183 arquivos e errou as classes de 3817 arquivos da base de dados.

TABELA VIII: Resultado da matriz de confusão da comparação dos resultados do índice MOS das execuções dos algoritmos ITU-T P.563 e ITU-T P.862.

	C1-P.862	C2-P.862	C3-P.862	C4-P.862
C1-P.563	4609	2381	350	104
C2-P.563	11	1559	35	4
C3-P.563	0	48	899	0
C4-P.563	0	0	0	0

Ao comparar os resultados alcançados do algoritmo P.563 com os resultados do obtidos pelo modelo, sendo que o modelo

foi treinado e validado de acordo com os resultados do MOS obtido pelo algoritmo da recomendação ITU-T P.862, podemos verificar que o modelo obteve um melhor desempenho que o algoritmo P.563, onde foi obtido uma acurácia de 96,467% para o modelo e de 61,83% para o algoritmo P.563. Deve-se destacar que o P.563 e algoritmo proposto são não intrusivos, precisando só de um sinal para determinação do índice MOS

## V. CONCLUSÃO

Através dos resultados alcançados pode-se concluir que o modelo proposto atende o objetivo principal do projeto que é classificar a taxa de perda de pacotes e identificar o índice MOS através de um algoritmo Deep Learning, no qual a sua taxa de acerto foi em média de 96,671% para o modelo treinado pela taxa de perda de pacotes e de 97,500% para o modelo treinado pelo índice MOS dos arquivos degradados de áudio. Podemos destacar que o modelo proposto obteve uma maior eficiência que o algoritmo da recomendação ITU-T P.563, sendo que ambos não utilizam um sinal de referência. Assim, a nossa proposta se torna um modelo de análise da qualidade de voz de forma objetiva e não intrusiva com maior eficiência, podendo através dele monitorar as transmissões VoIP reconhecendo a taxa de perda de pacotes em tempo real e possibilitando a execução de alguma ferramenta de correção para que a experiência do usuário não seja afetada.

## AGRADECIMENTOS

Os autores agradecem ao Setor de Extensão, Inovação, Pesquisa e Pós-Graduação (CFO- SEIPPG) do IFMG-Campus Formiga, pelo apoio financeiro.

## REFERÊNCIAS

- [1] I.-T. R. P.800, "Methods for subjective determination of transmission quality," Aug. 1996. [Online]. Available: <http://www.itu.int/rec/T-REC-P.800/en>
- [2] I.-T. R. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," 2001. [Online]. Available: <http://www.itu.int/rec/T-REC-P.862/en>
- [3] I.-T. R. P.563, "Single-ended method for objective speech quality assessment in narrow-band telephone applications," Apr. 2004. [Online]. Available: <http://www.itu.int/rec/T-REC-P.563/en>
- [4] H. Shu, Y. Song, and H. Zhou, "Time-frequency performance study on urban sound classification with convolutional neural network," in *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018, pp. 1713–1717.
- [5] G. Mittag and S. Möller, "Non-intrusive estimation of packet loss rates in speech communication systems using convolutional neural networks," in *2018 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2018, pp. 105–109.
- [6] E. T. Affonso, R. D. Nunes, R. L. Rosa, G. F. Pivaro, and D. Z. Rodríguez, "Speech quality assessment in wireless voip communication using deep belief network," *IEEE Access*, vol. 6, pp. 77 022–77 032, 2018.
- [7] J. Ooster and B. T. Meyer, "Improving deep models of speech quality prediction through voice activity detection and entropy-based measures," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 636–640.
- [8] E. T. Affonso, R. L. Rosa, and D. Z. Rodríguez, "Speech quality assessment over lossy transmission channels using deep belief networks," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 70–74, 2018.
- [9] J. Xu and C. Zhang, "Research of an improved packet loss compensation algorithm in voip," in *2011 IEEE 3rd International Conference on Communication Software and Networks*. IEEE, 2011, pp. 12–15.
- [10] I.-T. R. G.1050, "Network model for evaluating multimedia transmission performance over internet protocol," Nov. 2007. [Online]. Available: <http://www.itu.int/rec/T-REC-G.1050/en>