Accurate Analysis of Voice Disorder Using ResNet-50 Algorithm in Comparison with ResNet-18 Algorithm

Aakash S. S. and Bharatha Devi N.
Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

Keywords: Convolutional Neural Network, Health, Novel ResNet-50, ResNet-18, Speech, Voice Disorder.

Abstract: The study aims to enhance voice disorder detection precision using the novel ResNet-50 algorithm and comparing its efficacy with the ResNet-18 algorithm. for evaluating the accuracy of voice disorder identification, the research uses a confidence level of 95% and a g power of 0.8.w two algorithms, novel ResNet-50 and ResNet-18, are applied to a dataset of 864,448 mp3 audio files with accompanying metadata. the findings reveal that the novel ResNet-50 algorithm boasts an accuracy of 88.70%, superior to the 70.81% achieved by the ResNet-18 algorithm. however, with a significance value of 0.18 (independent sample t-test $p<0.05$), no noteworthy statistical difference was found between the two. in essence, the novel ResNet-50 algorithm demonstrates a higher accuracy in voice disorder analysis compared to the ResNet-18 algorithm.

1 INTRODUCTION

In this paper, the amalgamation of speech processing and machine learning techniques is utilised to discern disordered speech and subsequently categorise it as resulting from Neoplasm, Phonotrauma, or Vocal Palsy (Bhat and Kopparapu 2018). The research employs the Modified Mellin Transform of Log Spectrum (MMTLS) feature, derived to identify anomalous speech samples (Francis, Nair, and Radhika 2016). The aim is to introduce an automated method capable of distinguishing between healthy and pathological human voices in real-time. This facilitates more precise medical evaluations and encourages individuals with potential illnesses to pursue timely medical intervention (Milani, Ramashini, and Krishani 2020). As some individuals struggle with voice rhythm, voice recordings are often preferred over typing Arabic numerals. One practical application of this research is developing a dependable model to differentiate between normal, neoplastic, phototraumatic, and voice paralysed samples in the FEMH dataset (Al-Nasheri et al. 2018).

2 LITERATURE SURVEY

Recent research has focused on enhancing the identification of voice pathology using the Novel ResNet-50 algorithm. With 711 papers on IEEE Xplore and 92 articles in Sciencedirect, there is evident interest in the field. This study aims to offer physicians and logopaedicians fresh, objective metrics and illustrations to assess voice quality post-vocal fold surgery (Manfredi and Peretti 2006; Firdos and Umarani 2016; G. Ramkumar et al 2022). Its ambition is to craft a technique that differentiates between healthy and afflicted vocal patterns, leveraging a user-friendly approach (Wahed 2014). Furthermore, to gauge the recovery of a patient's vocal health following vocal fold surgery, this work introduces novel, easy-to-understand metrics and visuals for clinicians and logopedists (Manfredi and Peretti 2006; Umapathy et al. 2005; Padma, S et al. 2022).

A recognised gap in current research is the inadequate accuracy associated with present methods. Current techniques are hindered by limitations like the need for large datasets to predict accurately. Contrarily, this study's recommended Novel ResNet-50 algorithm achieves heightened accuracy by optimally utilising a smaller dataset for both training and validation. This research endeavours to elevate the vocal disorder detection efficacy of the Novel ResNet-50 approach.

3 PROPOSED METHODOLOGY

The research took place at the Image Processing laboratory within the Department of Computer

S., A. and N., B.
Accurate Analysis of Voice Disorder Using ResNet-50 Algorithm in Comparison with ResNet-18 Algorithm.
DOI: 10.5220/0012564600003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 502-507
ISBN: 978-989-758-661-3

Science and Engineering at Saveetha School of Engineering, a part of Saveetha Institute of Medical and Technical Sciences, Chennai. To determine the sample size, ClinCalc online software was employed, comparing both controllers. Two distinct groups were selected for comparative analysis. The study incorporated 40 samples in total, equally divided with 20 samples from each group (Borsky et al. 2017). Both the Novel ResNet-50 and ResNet-18 algorithms were applied using technical analysis software. Calculations were executed with 80% G-power, an alpha level of 0.05, a beta level of 0.2, all within a confidence interval of 95% [source: (https://clincalc.com/stats/samplesize.aspx)].

The Voice dataset consists of speech data extracted from public domain resources, such as user-submitted blog entries, historical books, classic films, and other spoken word collections read by Common Voice participants. Primarily aimed at aiding the development and testing of automatic speech recognition (ASR) systems, this dataset boasts 864,448 MP3 audio files. Accompanying metadata includes filenames, uttered phrases, regional accents, age, gender, and user feedback. This information is cataloged in the dataset's TSV files.

This study was conceptualised and actualised using Google Collab Python OpenCV software and was tested on the Windows 10 platform. The dataset from the Kaggle website (Zhang et al. 2020) aided in the code implementation. The hardware configuration consisted of an Intel Core i7 processor, 4GB RAM, and a 64-bit system architecture, with Python being the chosen programming language. The dataset was processed concurrently during code execution, culminating in an output detailing accuracy results.

3.1 Novel ResNet-50

The Residual Network, commonly referred to as ResNet, is a unique type of convolutional neural network (CNN) developed by He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian. CNNs have found extensive use in numerous computer vision applications. Among its variants, the ResNet-50 is illustrative, chosen for the initial sample grouping. Comprising 50 layers, the Novel ResNet-50 encompasses 48 convolutional layers, coupled with one MaxPool layer and one average pool layer. These networks are built by layering residual blocks. Originally, the design of the Novel ResNet-50 drew inspiration from ResNet-34, which consisted of 34 weighted layers. What sets Novel ResNet-50 apart is its pioneering method of integrating additional convolutional layers into a CNN without falling prey to the vanishing gradient problem. This is achieved via the introduction of shortcut connections.

Table 1: Procedure of the Novel ResNet-50 Algorithm.

Data Input: a training set with F features and n trees.
1. Provide initial values to the input variables.
2. From the available features list, choose the top k traits.
3. After finding the split point, precisely divide the dataset into child nodes.
4. Determine the decision tree's origin using the k attributes you've chosen.
5. Save the result (accuracy) using the test features and the decision trees that were plotted.
6. Collect the voting results for each conceivable reserved outcome and determine which outcome is most likely using this information.

Table 2: Procedure of the ResNet-18 Algorithm.

Input: Set of Exercises for Training Input
1. Give the input parameters as initial values.
2. Group the labels in the dataset into distinct categories.
3. For every attribute, probabilities and frequencies are determined.
4. The Naive Bayes model is used to calculate the likelihoods that follow from the features.
5. When all the probabilities have been estimated, every feature is multiplied by each probability.
6. Data are compared before being partitioned into groups.

To provide context, a 34-layer ResNet clocks in at 3.6 billion FLOPs, while its 18-layer counterpart operates at 1.8 billion FLOPs. This is substantially more efficient than a VGG-19 Network, which operates at a hefty 19.6 billion FLOPs. The intricacies of the Novel ResNet-50 algorithm are elaborated upon in Table 1.

3.2 ResNet-18

The ResNet-18 algorithm is utilised within the second sample preparation group. ResNet-18 is an 18-layer convolutional neural network, designed specifically to ensure the efficient operation of extensive convolutional neural network layers. Its architectural design is geared towards tackling the dilemma of sustaining performance amidst deepening networks. While deeper layers frequently culminate in deteriorating output quality, ResNet-18 seeks to counter this setback. The network houses close to 11 million trainable parameters and is structured with CONV layers and 3x3 filters, mirroring the VGG Net configuration. Only two pooling layers are interspersed within this network: one positioned at the beginning and the other towards the end. Each pair of CONV layers maintain identity relationships. An already trained version of ResNet-18 is accessible within the ImageNet database, having been educated on a dataset spanning more than a million images. This fine-tuned network boasts the prowess to categorise images across 1000 unique object categories, a spectrum that includes entities ranging from animals to keyboards and pencils. As a result, the network is adept at forming strong feature representations for a vast array of images. The network processes images at a resolution of 224 by 224 pixels. A comprehensive breakdown of the ResNet-18 methodology can be found in Table 2.

4 STATISTICAL ANALYSIS

The statistics for Novel ResNet-50 and ResNet-18 are evaluated using the SPSS software. The independent variables in this analysis include image, length, pitch, frequency, modulation, amplitude, volume, and decibels. Meanwhile, the dependent variables consist of pitch and volume. To ascertain the accuracy of both methods, a distinct T-test analysis is employed.

5 RESULTS

Twenty individuals were selected as a sample size for the execution of the Novel ResNet-50 and ResNet-18 algorithms using Anaconda Navigator. The subsequent comparative examination highlighted that the Novel ResNet-50 algorithm demonstrated superior accuracy in diagnosing voice abnormalities in comparison to the ResNet-18 algorithm.

Table 1 elucidates the operational procedure associated with the Novel ResNet-50 Algorithm. This model is characterised by a composition of 48 convolutional neural network layers, interspersed with one max pool layer and one average pool layer.

Table 2 delineates the workings of ResNet-18, an 18-layer convolutional neural network.

Table 3 presents a summary of the statistical analysis for both the Novel ResNet-50 and ResNet-18 algorithms, drawing from a dataset encompassing 20 samples. This table highlights the calculated mean values, standard deviations, and standard error means. The comparison between the Novel ResNet-50 and the ResNet-18 reveals a conspicuous advantage of the former in terms of mean accuracy and a lower mean loss.

Table 4 depicts the results derived from the Independent Sample T-test. With a significance value computed at 0.18 (considering an Independent Sample T-test $p<0.05$), the data suggests that there isn't a statistically significant difference discerned between the two examined groups.

Lastly, Figure 1 offers a visual representation in the form of a bar chart, contrasting the mean accuracy and loss metrics of the Novel ResNet-50 and ResNet-18 algorithms. The Novel ResNet-50's mean accuracy is visibly higher than that of its ResNet-18 counterpart.

Table 3: Group Statistical Analysis of Novel ResNet-50 and ResNet -18.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Novel ResNet-50	20	88.70	0.92850	0.20762
	ResNet -18	20	70.81	1.17558	0.26287
Loss	Novel ResNet-50	20	11.30	1.53756	0.34381
	ResNet -18	20	29.19	0.97023	0.21695

Table 4: Independent Sample T-test: The significant value obtained is p= 0.18 (Independent sample T-test p<0.05) which shows that there is no statistically significant difference between the two groups.

		Levene's test for equality of variances		T-test for equality means with 95% confidence interval						
		f	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. Error difference	Lower	Upper
Accuracy	Equal variances assumed	0.770	0.386	53.408	38	0.18	17.8900	0.3349	17.211	18.5681
	Equal Variances not assumed			53.408	36.064	0.18	17.8900	0.3349	17.21	18.5693
Loss	Equal variances assumed	1.397	0.245	-44.043	38	0.08	-17.9050	0.40654	-18.727	-17.0820
	Equal Variances not assumed			-44.043	32.060	0.08	-17.90500	0.40654	-18.73303	-17.07697

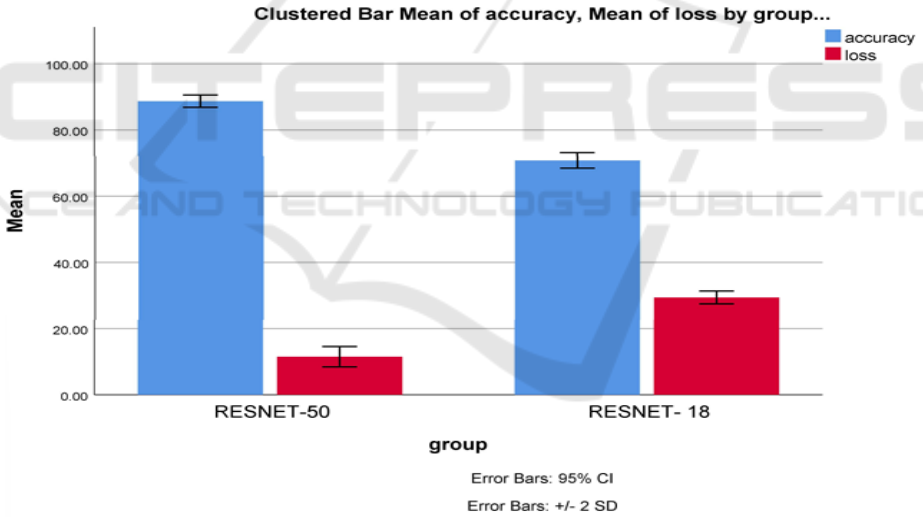

Figure 1: Comparison of Novel ResNet-50 and ResNet -18. Classifier in terms of mean accuracy and loss. The mean accuracy of Novel ResNet-50 is better than ResNet -18. X-Axis: Novel ResNet-50 Vs ResNet -18 Classifier, Y-Axis: Mean accuracy: Error Bar +/- 2SD.

6 DISCUSSION

The Novel ResNet-50 boasts an accuracy of 88.70%, outstripping the ResNet-18, which stands at 70.81%. With the study's significance pegged at 0.18 (using an Independent Sample T-test with p<0.05), it insinuates the superiority of the Novel ResNet-50 over the ResNet-18.

The main thrust of this paper pivots on the exploration and juxtaposition of various machine learning strategies applied in the detection of voice disorders. The research intimates that depending on the attributes evaluated via apt feature selection techniques, either the decision tree algorithm or the support vector machine algorithm notches up an accuracy rate of 84.3% (Verde, De Pietro, and

Sannino 2018). The study's lens is trained on a spectrum of acoustic characteristics derived from vocal fold signals, chiefly zeroing in on pitch. Experimentally, the earmarked features have been adjudged to be of immense import, as the classification algorithm, even in its unadorned form, touches an apex accuracy rate of 91.5% (Umapathy et al. 2005). The VGG-16 CNN model, together with the Convolutional Neural Network, have been utilised in this endeavour. The experiment exploited hundreds of PVD audio files from the Respiratory Sound Database, exploring the CNN's prowess in pinpointing aberrant speech. The diagnosis of voice pathology was discerned with a precision of 92.03% (Gumelar et al. 2020). The overarching aim of this scrutiny is to evaluate and draw parallels between machine learning methods tailored for the precocious detection of Voice Disorders, even before the symptoms unfurl. The proposed paradigm has been validated to clock a staggering 93% accuracy in the allotted endeavour, employing a conglomerate of learning models (Hussain and Sharma 2022).

The research methodology wends its way through data amassed from variegated reservoirs, contending with the challenge of voice data recognition. Yet, the study doesn't emerge unscathed from constraints; a conspicuous drawback is the protracted span earmarked for dataset training. Envisioning the road ahead, the research aspires to amplify the system's ambit, embracing an enlarged cadre of subjects, whilst concurrently curtailing the duration expended on dataset training.

7 CONCLUSION

Voice disorders, often neglected in the broader spectrum of medical issues, are essential for diagnostics, given the significant role voice plays in human communication. The advanced machine learning algorithms we've discussed in this study, especially the Novel ResNet-50 and ResNet-18, have the potential to revolutionize this area of diagnosis. The insights derived from our comparison not only spotlight the competencies of these algorithms but also delineate the path ahead for further exploration. Summarizing the findings, we can highlight six cardinal points:

- Depth of Algorithm: The layer configuration in the Novel ResNet-50, with its 50 layers, provides a depth that seems conducive to intricate voice analysis, besting the shallower ResNet-18.
- Handling Vanishing Gradient: The ingenuity of the Novel ResNet-50 resides in its inventive approach of adding more convolutional layers without facing the vanishing gradient problem, constraint often limiting deep neural networks.
- Pre-trained Networks: The availability of pretrained versions, especially for ResNet-18 on extensive databases like ImageNet, indicates their potential adaptability to diverse tasks, including voice disorder detection.
- Feature Representation: The networks' ability to categorise and represent a multitude of features ensures that they capture the intricacies of voice patterns, making the diagnosis precise and accurate.
- Training Time: One trade-off for the increased accuracy observed in Novel ResNet-50 could be the training time. As the layers increase, so does the computation demand, an area where ResNet-18 might have an advantage.
- Future Applications: Given the efficacy of the Novel ResNet-50 in voice disorder detection, it offers promising prospects in other domains requiring meticulous pattern recognition.

In conclusion, this study pivots around the comparative analysis of the Novel ResNet-50 and ResNet-18 in the context of voice disorder detection. Evidently, the Novel ResNet-50, with an accuracy metric of 88.70%, outshines the ResNet-18, which clocks an accuracy of 70.81%. This differential underscores the robustness and superiority of the Novel ResNet-50 paradigm over its ResNet-18 counterpart. The comprehensive exploration furnished in this study not only underscores the inherent strengths and limitations of each algorithm but also offers a clarion call to researchers to further delve into this promising arena.

REFERENCES

Al-Nasheri, Ahmed, Ghulam Muhammad, Mansour Alsulaiman, Zulfiqar Ali, Khalid H. Malki, Tamer A. Mesallam, and Mohamed Farahat Ibrahim. 2018. “Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions.” *IEEE Access* 6: 6961–74.

Alvarez, Mauricio, Ricardo Henao, Germán Castellanos, Juan I. Godino, and Alvaro Orozco. 2006. “Kernel Principal Component Analysis through Time for Voice Disorder Classification.” *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE*

Engineering in Medicine and Biology Society. Conference 2006: 5511–14.

Bhat, Chitralekha, and Sunil Kumar Kopparapu. 2018. "FEMH Voice Data Challenge: Voice Disorder Detection and Classification Using Acoustic Descriptors." In *2018 IEEE International Conference on Big Data (Big Data)*, 5233–37.

Eskidere, Ömer, Ömer Aktaş, and Cevat Ünal. 2015. "Voice Disorders Identification Using Discrete Wavelet Based Features." In *2015 Medical Technologies National Conference (TIPTEKNO)*, 1–4.

Firdos, Seema, and K. Umarani. 2016. "Disordered Voice Classification Using SVM and Feature Selection Using GA." In *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*, 1–6.

Francis, Christina Raichel, Vrinda V. Nair, and Salini Radhika. 2016. "A Scale Invariant Technique for Detection of Voice Disorders Using Modified Mellin Transform." In *2016 International Conference on Emerging Technological Trends (ICETT)*, 1–6.

Goldstein, Anatoly D., and Robert E. Hillman. 2012. "Integration, Reuse and Sharing of Data on Voice Disorders." In *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*, 407–14.

G. Ramkumar, G. Anitha, P. Nirmala, S. Ramesh and M. Tamilselvi, "An Effective Copyright Management Principle using Intelligent Wavelet Transformation based Water marking Scheme," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-7, doi: 10.1109/ACCAI53970.2022.9752516.

Gumelar, Agustinus Bimo, Eko Mulyanto Yuniarno, Wiwik Anggraeni, Indar Sugiarto, Vincentius Raki Mahindara, and Mauridhi Hery Purnomo. 2020. "Enhancing Detection of Pathological Voice Disorder Based on Deep VGG-16 CNN." *2020 3rd International Conference on Biomedical Engineering (IBIOMED)*. https://doi.org/10.1109/ibiomed50285.2020.9487589.

Hussain, Audil, and Amit Sharma. 2022. "Machine Learning Techniques for Voice-Based Early Detection of Parkinson's Disease." *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. https://doi.org/10.1109/icacite53722.2022.9823467.

Padma, S., Vidhya Lakshmi, S., Prakash, R., Srividhya, S., Sivakumar, A. A., Divyah, N., ... & Saavedra Flores, E. I. (2022). Simulation of land use/land cover dynamics using Google Earth data and QGIS: a case study on outer ring road, Southern India. Sustainability, 14(24), 16373

Manfredi, Claudia, and Giorgio Peretti. 2006. "A New Insight into Postsurgical Objective Voice Quality Evaluation: Application to Thyroplastic Medialization." *IEEE Transactions on Bio-Medical Engineering* 53 (3): 442–51.

Milani, M. G. Manisha, Murugaiya Ramashini, and Murugiah Krishani. 2020. "A Real-Time Application to Detect Human Voice Disorders." In *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 979–84.

Umapathy, Karthikeyan, Sridhar Krishnan, Vijay Parsa, and Donald G. Jamieson. 2005. "Discrimination of Pathological Voices Using a Time-Frequency Approach." *IEEE Transactions on Bio-Medical Engineering* 52 (3): 421–30.

Verde, Laura, Giuseppe De Pietro, and Giovanna Sannino. 2018. "Voice Disorder Identification by Using Machine Learning Techniques." *IEEE Access*. https://doi.org/10.1109/access.2018.2816338.

Wahed, Manal Abdel. 2014. "Computer Aided Recognition of Pathological Voice." *2014 31st National Radio Science Conference (NRSC)*. https://doi.org/10.1109/nrsc.2014.6835096.

Zhang, Tao, Yangyang Shao, Yaqin Wu, Zhibo Pang, and Ganjun Liu. 2020. "Multiple Vowels Repair Based on Pitch Extraction and Line Spectrum Pair Feature for Voice Disorder." *IEEE Journal of Biomedical and Health Informatics* 24 (7): 1940–51.