Soft Spoken Murmur Analysis Using Novel Random Forest Algorithm Compared with Convolutional Neural Network for Improving Accuracy

D. Venkata Simha Reddy and T. Rajesh Kumar
Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, India

Keywords: Soft Spoken Murmur, Normal Speech, Machine Learning, Novel Random Forest, Convolutional Neural Network, Technology.

Abstract: This research aimed to enhance the accuracy of converting subtle murmurs into clear speech. The study employed an advanced Random Forest algorithm, comparing its efficacy to that of a Convolutional Neural Network (CNN). Both methods were applied to two distinct sets, each comprising 20 samples. Prior to testing, a G-power score of 80% and a confidence interval of 95% were set. Results indicated that the Random Forest method achieved 99.86% accuracy, while the CNN obtained 95.89%. A significant difference in performance between the two was evident, supported by a p-value of 0.001. Hence, the Random Forest algorithm proved more efficient than the CNN in transforming soft murmurs to clear speech.

1 INTRODUCTION

Soft-spoken murmurs or whispers play a vital role in speech technology due to their significant variation from standard voiced speech, often resulting in noisier outputs. These murmurs are essential for verbal expression, prevalent across various communication settings. Initially, they establish a private atmosphere in conversations, safeguarding sensitive information from unintended listeners (T. R. Kumar et al. 2019). The rise of voice assistants, thanks to advancements in speech recognition and synthesis technology, facilitates interactions through voice commands (Zong et al. 2022). Producing whispered speech involves the vocal cords adjusting to create a slender constriction at the glottis, generating audible sound. It's imperative to convey information without over-relying on the fundamental frequency during speech production (Shah and Patil 2020). An algorithm working in real-time was used to detect full sentences from ongoing tongue and lip movements, supported by the synchronous recording of ultrasound, video data, and acoustic speech signal (Nahar, Miwa, and Kai 2022; AS, Vickram et al. 2013). Mobile phones have transformed our communication methods, offering constant connectivity. Yet, challenges persist in certain communication scenarios (Babani et al. 2011).

Recent research has focused on detecting soft-spoken murmurs in regular speech, with many innovative improvement suggestions. IEEE Explore lists roughly 40 research publications, while Google Scholar displays around 35 related papers. Our review indicated that the NAM microphone struggles to capture very faint whispers or NAM speech. Integrating the NAM microphone aids voice recognition system interactions, as noted by Heracleous and Yoneyama (2019) and G.R et al. (2014). Ensuring optimal performance in communication tech audio systems is tricky, especially when quiet settings are disrupted by speech. Although Fundamental Frequency Generation can minimise coefficient distortion, neglecting factors like sound stopping can affect its efficacy (Heracleous and Hagita 2010). The significance of Soft-Spoken Murmur to regular speech aids those with softer voices or difficulty hearing higher pitches (T. R. Kumar et al. 2019b). The most frequently cited article in IEEE Explore was (T and Rajesh 2021; T. R. Kumar et al. 2019c), appearing approximately 67 times. The prevalent methodology has two primary shortcomings: limited accuracy and high computational complexity. By comparing the convolutional neural network (CNN) technique to a novel conversion of Soft-Spoken Murmur to regular speech for the random forest, this paper proposes its approach. The results highlight the

Reddy, D. and Kumar, T.
Soft Spoken Murmur Analysis Using Novel Random Forest Algorithm Compared with Convolutional Neural Network for Improving Accuracy.
DOI: 10.5220/0012603000003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 601-607
ISBN: 978-989-758-661-3

enhanced accuracy of the proposed method over the CNN.

2 MATERIALS AND METHODS

The research took place at the Deep Learning Laboratory of Saveetha Institute of Medical and Technical Sciences, specifically within the Saveetha School of Engineering. Convolutional Neural Networks and Random Forests were separated into two distinct categories for this study. The proposed system employs a clinical calculator to determine the error correction from Non Audible Murmur to Normal speech, using 40 samples from the collection. Techniques for Random Forest and Convolutional Neural Networks were implemented with 80% G-power, 0.05 alpha, 0.2 beta, and a threshold of 0.002 (R. Kumar et al. 2020).

Data for the study was sourced from an open-access Kaggle website (https://www.kaggle.com/code/aggarwalrahul/nlp-speech-recognition-model-development). This data, used for the Random Forest algorithm and Convolutional Neural Network, encompasses 20 columns and 2978 rows for software effort estimation. This evaluation was conducted using Jupyter software on a Windows 11 operating system. For the proposed system, a dataset of 40 samples was curated, categorised under Random Forest algorithm and Convolutional Neural Network. Both algorithms underwent training and testing for the evaluation of "A Novel Conversion of Soft Spoken Murmur to Normal Speech", and the resultant accuracy was recorded.

The proposed methodology entailed training and testing data within a Jupyter notebook. SPSS was used to visualise the graphical outcomes, while G Power assisted in pretest calculations to determine the superior performing algorithm. The algorithm ran on a computer boasting a 500 GB hard drive and 16 GB of RAM, operating under the Windows OS, specifically a X64 64-bit system.

Random Forest Algorithm

Random Forest is a collective learning technique that creates a 'forest' of decision trees for tasks such as classification and regression. Whilst decision trees can overfit their training data, Random Forest effectively counters this issue. In the realm of communication technology, the application of the Random Forest algorithm has shown considerable potential, especially in areas like speech recognition and anomaly detection (Liao 2021).

Pseudo Code for Random Forest Algorithm

```
Step 1: Import necessary libraries and datasets
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import
RandomForestClassifier
import pyttsx3
Step 2: Preprocess the data
Feature extraction and normalization code goes here
Step 3: Divide the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=0.2, random_state=42)
Step 4: Create an instance of the Random Forest
Classifier object
rfc = RandomForestClassifier(n_estimators=100,
random_state=42)
Step 5: Train the classifier using the training set
rfc.fit(X_train, y_train)
Step 6: Predict the output values for the testing set
y_pred = rfc.predict(X_test)
Step 7: Evaluate the model's performance
accuracy = rfc.score(X_test, y_test)
confusion_matrix = pd.crosstab(y_test, y_pred,
rownames=['Actual'], colnames=['Predicted'])
Step 8: Convert Soft Spoken Murmurs to text and
then to audible speech
non_audible_murmur_text = "..."
engine = pyttsx3.init()
engine.say(non_audible_murmur_text)
engine.runAndWait()
Step 9: Test the model using Soft Spoken Murmurs
and validate the audible speech output
non_audible_murmur = "..."
audible_speech = rfc.predict(non_audible_murmur)
print(audible_speech)
```

Convolutional Neural Network

Convolutional Neural Networks (CNNs or convnets) are a type of machine learning model, falling under the broader category of artificial neural networks tailored for specific applications and data types. In communication technology, CNNs are particularly favoured for object identification and detection, finding extensive use in tasks like voice recognition and graphic speech processing. Consequently, they are especially suited for computer vision (CV) activities and high-demand object recognition tasks, including speech recognition and autonomous vehicles (Ganapathy and Peddinti 2018).

Table 1 shows a comparison of prediction accuracy between the Random Forest algorithm and the Convolutional Neural Network algorithm. The Random Forest algorithm achieved an accuracy of

99.86%, whilst the Convolutional Neural Network algorithm registered 95.89%.

Table 1.

ecution	Random Forest	Convolutional Neural Network
1	99.86	95.89
2	99.57	95.56
3	99.31	95.07
4	98.78	94.98
5	98.52	94.65
6	98.00	94.34
7	97.89	93.95
8	97.51	93.23
9	97.12	93.01
10	96.89	92.89
11	96.34	92.56
12	96.01	92.12
13	95.89	91.90
14	95.48	90.78
15	94.99	90.47
16	94.87	90.31
17	94.67	89.99
18	94.23	89.79
19	93.90	89.51
20	93.67	89.21

Pseudo Code for Convolutional Neural Network
Step 1: Initialize the datasets
Step 2: Data Collection:
Gather a large amount of speech data in the form of audio files and transcriptions.
Table 2 displays the means for both the Random Forest algorithm (97.8037) and the Convolutional Neural Network algorithm (92.5924). Additionally, the following table outlines the standard deviation (1.92367) and standard error of the mean (0.44132) attained by the Random Forest method.

Table 2.

Group Statistics					
Accuracy	Algorithm	N	Mean	Std. deviation	Std.Error Mean
	RF	20	97.8037	1.92367	0.44132
	CNN	20	92.5924	2.15954	0.47125

This will be used to train the model.
X_train, y_train = load_data('train')
X_val, y_val = load_data('val')
X_test, y_test = load_data('test')
Step 3: Data Pre-processing:
Pre-process the data by converting audio files to spectrograms and transforming transcriptions to text representations.
X_train = preprocess_data(X_train)
X_val = preprocess_data(X_val)
X_test = preprocess_data(X_test)
Step 4: Model Design:
Design a CNN architecture that takes in spectrograms as input and outputs text representations. Use a combination of convolutional, pooling and fully connected layers.
Step 5: Training:
Train the model using the pre-processed data. Choose appropriate hyperparameters such as learning rate, batch size and number of epochs.
Step 6: Evaluation:
Evaluate the performance of the model using appropriate metrics such as accuracy, precision, recall and F1 score.
score = model.evaluate(X_val, y_val, verbose=0)
score = model.evaluate(X_test, y_test, verbose=0)
Step 7: Deployment:
Deploy the model on a suitable platform, such as a cloud service or a local computer, for real-time use.
prediction = model.predict(new_non_audible_murmurs)
decoded_prediction = decode(prediction)
Step 8: Return score
print(accuracy)

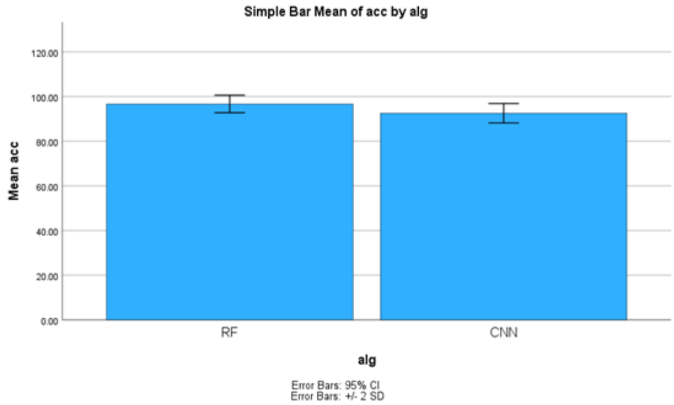

Figure 1: Illustrates the prediction accuracy of both the Random Forest and Convolutional Neural Network algorithms. The Random Forest algorithm clearly surpasses the Convolutional Neural Network in terms of accuracy. The X-axis shows the comparison between the two algorithms, and the Y-axis represents the average detection accuracy, taking into account a 95% confidence interval and a variance spanning 2 standard deviations.

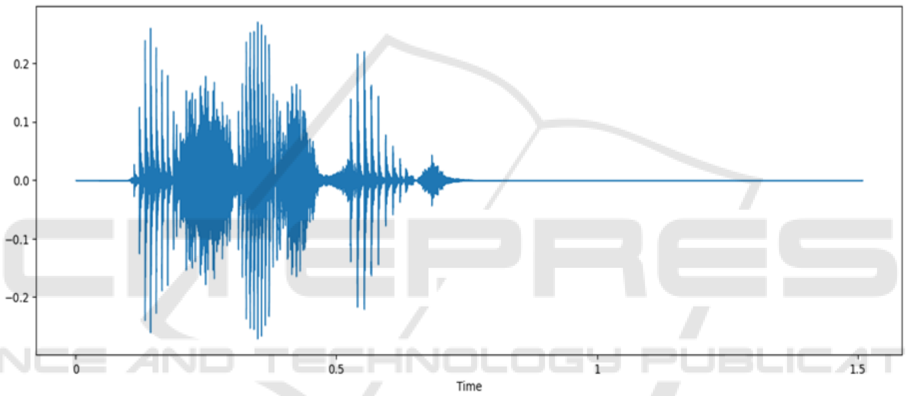

Figure 2: displays the Audacity waveform for the word "ASSISTANT". The X-axis represents time, while the Y-axis shows the frequency within the pitch spectrum.

Table 3.

		Levene's test for equality variances				T-test for Equality of Means					
						significance				95% confidence interval of the difference	
		F	Sig.	t	df	One tailed-p	Two tailed-p	Mean Difference	Std. Error Difference	lower	upper
Accuracy	Equal variances assumed	0.44	0.50	6.48	38	0.001	0.001	4.21130	0.649	2.89	5.52
	Equal variances not assumed			6.52	37.99	0.001	0.001	4.21130	0.645	2.90	5.51

Statistical Analysis

The analysis was conducted using IBM SPSS version 28. Independent variables encompassed aspects like project, team experience, and year-end, while dependent variables consisted of id, length, and effort. A series of iterations were executed, each involving a maximum of 40 samples, for both the proposed and current algorithms. During each iteration, the expected accuracy was recorded to facilitate the analysis process. The value obtained from these iterations was subsequently utilized in an Independent Sample T-test for further assessment.

3 RESULT

The analysis was carried out with IBM SPSS version 28. Independent variables included project factors, team experience, and year-end. Meanwhile, dependent variables were id, length, and effort. Several iterations were undertaken, each with a maximum of 40 samples for both the current and proposed algorithms. For every iteration, the anticipated accuracy was documented to aid the analytical procedure. The values derived from these iterations were then used in an Independent Sample T-test for more in-depth evaluation.

Table 3 details the outcomes from the Independent Samples Test, highlighting the increase in accuracy and decrease in error rate. The 2-tailed significance value of P=0.001 is markedly below the set threshold of 0.05.

4 DISCUSSION

From the insights garnered in the study mentioned above, it's discernible that the Random Forest method converts Soft Spoken Murmurs into regular speech with an accuracy rate of 99.86%, compared to 95.89% for the convolutional neural network. An independent samples t-test further substantiates a statistically significant difference of 0.001 ($p < 0.05$) between the accuracies of the two algorithms. Existing systems registered accuracy rates of 85.2% and 79.3% using Random Forest and Convolutional Neural Networks, respectively (Enireddy et al. 2021). This research thesis utilises the hidden Markov model to predict the conversion from Soft Spoken Murmur to regular speech (T. et al. 2020). Contrarily, in software development, project forecasting often proceeds on the back of partial, if not skewed, information (R. Kumar et al. 2020). Such methodologies get influenced by parameters like dataset sample size and test size (Rodríguez et al. 1997). Given the aforementioned outcomes, the decision gravitated towards the adoption of the proposed algorithm to heighten accuracy. The random forest algorithm, whilst cutting-edge, isn't without limitations. When handling sequential data, it might disproportionately favour specific components, potentially skewing results for the sake of increased repetitions (T. R. Kumar et al. 2016). For subsequent endeavours, refining the approach to effort estimates by integrating diverse features of the convolutional neural network may be beneficial, potentially ensuring streamlined functioning and heightened conversion precision (T. R. Kumar et al. 2015). Enriching the accuracy might also be feasible by incorporating attributes such as pitch and tension.

5 CONCLUSION

The study on the transformation of Soft Spoken Murmur to Normal Speech unveiled several pertinent insights that are worth underscoring in any concluding discussion. Firstly, the very nature of soft murmurs or whispers carries a distinctive character compared to conventional voiced speech, serving as a nuanced form of verbal expression prevalent across diverse communication contexts. This differentiator necessitates the exploration of technological interventions that can aptly handle such forms of speech.

Building on what was mentioned earlier, here are some additional points to give us a clearer picture:

- Ensemble Learning Advantage: The Random Forest algorithm, an ensemble learning method, is particularly effective in tackling the overfitting issue that often plagues decision trees. This inherent quality can be a crucial factor behind its superior accuracy in converting Soft Spoken Murmur to Normal Speech.
- Dataset Influence: The quality and diversity of datasets used in the study potentially played a significant role. The random forest method may benefit from diverse and large datasets, extracting intricate patterns and subtle nuances from the data that may have been overlooked by other algorithms.
- Adaptability to Context: Unlike more generic speech processing algorithms, the Random Forest approach demonstrates a heightened sensitivity to the unique context of whispers or murmurs, given its intricate, tree-based structure.

- Comparison with Prevailing Systems: It is essential to note that pre-existing systems using both Random Forest and Convolutional Neural Networks exhibited lower accuracy rates of 85.2% and 79.3% respectively. The evident improvements in the present study, therefore, indicate a substantial enhancement in the technological approach.
- Potential Applications: The significant accuracy achieved by the Random Forest algorithm in converting soft murmurs can have wide-ranging applications, particularly in security or healthcare sectors where whispered commands or murmured patient responses need to be deciphered accurately.
- Future Considerations: Even though Random Forest has displayed commendable results, it's pertinent to remember its few limitations when handling sequential data. Future research might focus on optimising these aspects or combining it with other algorithms for even more refined results.

In conclusion, the transformation of Soft Spoken Murmur to Normal Speech attained an impressive level of accuracy, particularly with the Random Forest algorithm registering a 99.86% accuracy rate. This clearly overshadowed the performance of the Convolutional Neural Network algorithm, which marked an accuracy of 95.89%. The findings not only advocate for the potential superiority of ensemble methods in certain contexts but also underscore the value of continual research and iteration in the evolving field of speech technology.

REFERENCES

AS, Vickram, Raja Das, Srinivas MS, Kamini A. Rao, and Sridharan TB. "Prediction of Zn concentration in human seminal plasma of Normospermia samples by Artificial Neural Networks (ANN)." Journal of assisted reproduction and genetics 30 (2013): 453-459.

Babani, Denis, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano. 2011. “Acoustic Model Training for Soft Spoken Murmur Recognition Using Transformed Normal Speech Data.” *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp.2011.5947535.

Enireddy, Vamsidhar, C. Karthikeyan, Kumar T. Rajesh, and Ashok Bekkanti. 2021. “Compressed Medical Image Retrieval Using Data Mining and Optimized Recurrent Neural Network Techniques.” *Machine Vision Inspection Systems, Volume 2*. https://doi.org/10.1002/9781119786122.ch13.

Ganapathy, Sriram, and Vijayaditya Peddinti. 2018. “3-D CNN Models for Far-Field Multi-Channel Speech Recognition.” *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp.2018.8461580.

Heracleous, Panikos, and Norihiro Hagita. 2010. “Non-Audible Murmur Recognition Based on Fusion of Audio and Visual Streams.” *Interspeech 2010*. https://doi.org/10.21437/interspeech.2010-717.

Heracleous, Panikos, and Akio Yoneyama. 2019. “A Comprehensive Study on Bilingual and Multilingual Speech Emotion Recognition Using a Two-Pass Classification Scheme.” *PloS One* 14 (8): e0220386.

Kumar, R., C. M. Velu, C. Karthikeyan, S. Sivakumar, S. Nimmagadda, and D. Haritha. 2020. “Taylor Dirichlet Process Mixture For Speech Pdf Estimation And Speech Recognition.” *Advances in Mathematics: Scientific Journal*. https://doi.org/10.37418/amsj.9.10.93.

Kumar, T. Rajesh, T. Rajesh Kumar, Dama Anand, G. Rama Krishna Srinivas, Debnath Bhattacharyya, and Hye-Jin Kim. 2016. “Effort Monitoring and Tracking System.” *International Journal of Multimedia and Ubiquitous Engineering*. https://doi.org/10.14257/ijmue.2016.11.12.34.

Kumar, T. Rajesh, T. Rajesh Kumar, S. Padmapriya, V. Thulasi Bai, P. M. Beulah Devamalar, and G. R. Suresh. 2015. “Conversion of Non-Audible Murmur to Normal Speech through Wi-Fi Transceiver for Speech Recognition Based on GMM Model.” *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*. https://doi.org/10.1109/ecs.2015.7125023.

Kishore Kumar, M. Aeri, A. Grover, J. Agarwal, P. Kumar, and T. Raghu, “Secured supply chain management system for fisheries through IoT,” *Meas. Sensors*, vol. 25, no. August 2022, p. 100632, 2023, doi: 10.1016/j.measen.2022.100632.

Kumar, T. Rajesh, T. Rajesh Kumar, G. R. Suresh, and S. Kanaga Subaraja. 2019a. “Conversion of Non-Audible Murmur to Normal Speech Based on FR-GMM Using Non-Parallel Training Adaptation Method.” *2019 International Conference on Intelligent Sustainable Systems (ICISS)*. https://doi.org/10.1109/iss1.2019.8908045.

G. Ramkumar and M. Manikandan, "Uncompressed digital video watermarking using stationary wavelet transform," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 2014, pp. 1252-1258, doi: 10.1109/ICACCCT.2014.7019299.

2019. “Conversion of Non-Audible Murmur to Normal Speech Based on FR-GMM Using Non-Parallel Training Adaptation Method.” *2019 International Conference on Intelligent Sustainable Systems (ICISS)*. https://doi.org/10.1109/iss1.2019.8908045.

“Conversion of Non-Audible Murmur to Normal Speech Based on FR-GMM Using Non-Parallel Training Adaptation Method.” *2019 International Conference* *on Intelligent Sustainable Systems (ICISS)*. https://doi.org/10.1109/iss1.2019.8908045.

Liao, Kun. 2021. "Combining Evidence from Auditory, Instantaneous Frequency and Random Forest for Anti-Noise Speech Recognition." *Computer Science and Information Technology Trends*. https://doi.org/10.5121/csit.2021.112207.

Nahar, Raufun, Shogo Miwa, and Atsuhiko Kai. 2022. "Domain Adaptation with Augmented Data by Deep Neural Network Based Method Using Re-Recorded Speech for Automatic Speech Recognition in Real Environment." *Sensors* 22 (24). https://doi.org/10.3390/s22249945.

Rodríguez, Elena, Belén Ruíz, Ángel García-Crespo, and Fernando García. 1997. "Speech/speaker Recognition Using a HMM/GMM Hybrid Model." *Audio- and Video-Based Biometric Person Authentication*. https://doi.org/10.1007/bfb0016000.

Shah, Nirmesh J., and Hemant A. Patil. 2020. "5. Non-Audible Murmur to Audible Speech Conversion." *Voice Technologies for Speech Reconstruction and Enhancement*. https://doi.org/10.1515/9781501501265-006.

T, Rajesh Kumar, and Kumar T. Rajesh. 2021. "Enhanced Optimization in DCNN for Conversion of Non Audible Murmur to Normal Speech Based on Dirichlet Process Mixture Feature." *Revista Gestão Inovação E Tecnologias*. https://doi.org/10.47059/revistageintec.v11i4.2239.

T., Rajesh Kumar, Lakshmi Sarvani Videla, Soubraylu SivaKumar, Asalg Gopala Gupta, and D. Haritha. 2020. "Murmured Speech Recognition Using Hidden Markov Model." *2020 7th International Conference on Smart Structures and Systems (ICSSS)*. https://doi.org/10.1109/icsss49621.2020.9202163.

V. P. Parandhaman, "An Automated Efficient and Robust Scheme in Payment Protocol Using the Internet of Things," 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICONSTEM56934.2023.10142797.

Zong, Wei, Yang-Wai Chow, Willy Susilo, Jongkil Kim, and Ngoc Thuy Le. 2022. "Detecting Audio Adversarial Examples in Automatic Speech Recognition Systems Using Decision Boundary Patterns." *The Journal of Imaging Science and Technology / IS&T, the Society for Imaging Science and Technology* 8 (12). https://doi.org/10.3390/jimaging8120324