Science and Technology Publishing (SCI & TECH)
ISSN: 2632-1017
Vol. 9 Issue 3, March - 2025

Modelling Of Multilingual Speaker Recognition In Noisy Environment Using Random Forest Algorithm

Jimoh Jacob Afolayan[1]
Department of Electrical / Electronic Engineering
University of Uyo, Akwa Ibom State

Kingsley M. Udofia[2]
Department of Electrical / Electronic Engineering
University of Uyo, Akwa Ibom State

Kufre M. Udofia[3]
Department of Electrical / Electronic Engineering
University of Uyo, Akwa Ibom State

***Abstract*— Modelling of multilingual speaker recognition in noisy environment using Random Forest (RF) algorithm is presented. Basically, this research work presented RF algorithm speaker identification for a multilingual speech dataset consisting of speech samples from different Nigerian languages. The work compared the performance of the system when the RF model is trained using the clean speech dataset with no noise and when the RF model is trained with composite data that has some noise levels in the speech signal. Speech samples were collected from 15 different people where each of the speech samples lasted for a maximum of 120 seconds, with signal to noise ratio (SNR) ranging from 0 dB to 30 dB while the noiseless environment with high SNR is given a finite value of 100 dB. In practice it is assumed to be infinite. The results showed that the accuracy of the clean signal-trained model attained 85 % whereas the composite signal-trained model attained 96 %. Again, the percentage improvement in accuracy for using composite data trained model showed minimum improvement of 13 % and maximum improvement of 284 % over the cleaned signal trained model. Similar performance improvements were recorded for precision, F1_score and recall which showed that training the RF model with the composite dataset ensures that model superior performance is attained in all the noise levels the test was conducted.**

Keywords — Speaker Identification, Multilingual, Speaker Recognition, Noisy Signal, Random Forest Regression Algorithm

1. Introduction

Nigeria's linguistic diversity, with over 500 languages and numerous dialects shaped by ethnic, regional, and socio-economic factors, presents a unique challenge for speech-based technologies [1,2]. Global datasets often fail to represent this complexity, leading to underrepresentation of Nigerian languages and dialects in speaker identification and biometric systems. This research aims to address this gap by expanding the dataset to include a broader range of Nigerian languages and dialects, ensuring a more inclusive approach to speech-based biometric identification [3,4,5].

In addition, current speaker recognition systems often overlook the impact of environmental noise on performance [6,7]. There is a limited understanding of how varying levels of background noise, particularly across different signal-to-noise ratios (SNRs), influence system accuracy [8,9,10]. This research aims to explore the effects of noise on speaker identification performance, conduct experiments under various SNR conditions, identify the performance degradation thresholds, and develop strategies to enhance the robustness of speaker identification systems in noisy environments [11,12,13].

Specifically, this research work employed random forest algorithm speaker identification for a multilingual speech dataset consisting of speech samples from different Nigerian languages [14,15]. The work compared the performance of the system when the RF model is trained using the clean speech dataset with no noise and when the RF model is trained with composite data that has some noise levels in the speech signal. The details of the data acquisition, model training and evaluation are presented along with the results and discussions.

2. Methodology

This work utilized Random Forest (RF) machine learning model for speaker identification using multilingual speaker speech signals database captured in environment with different signal to noise ratio (SNR) levels. The clean signal is assumed to have an extremely high SNR while the noisy signal has very low value of SNR.

Basically, Random Forest (RF) is an ensemble learning method that builds multiple decision trees and aggregates their outputs to improve accuracy and reduce overfitting [17]. Each tree is trained on a bootstrap sample of the training data. The classification output is determined by majority voting. The RF model consists of m decision trees $T_1, T_2, \ldots, T_m$, and the final prediction $\hat{y}$ is made by taking the majority vote of the trees:

$$\hat{y} = mode(T_1(x), T_2(x), \ldots, T_m(x)) \quad (1)$$

where x is the input feature vector. The architecture of the Random Forest (RF) model is given in Figure 1 while the flow diagram of the model training and evaluation is given in Figure 2.

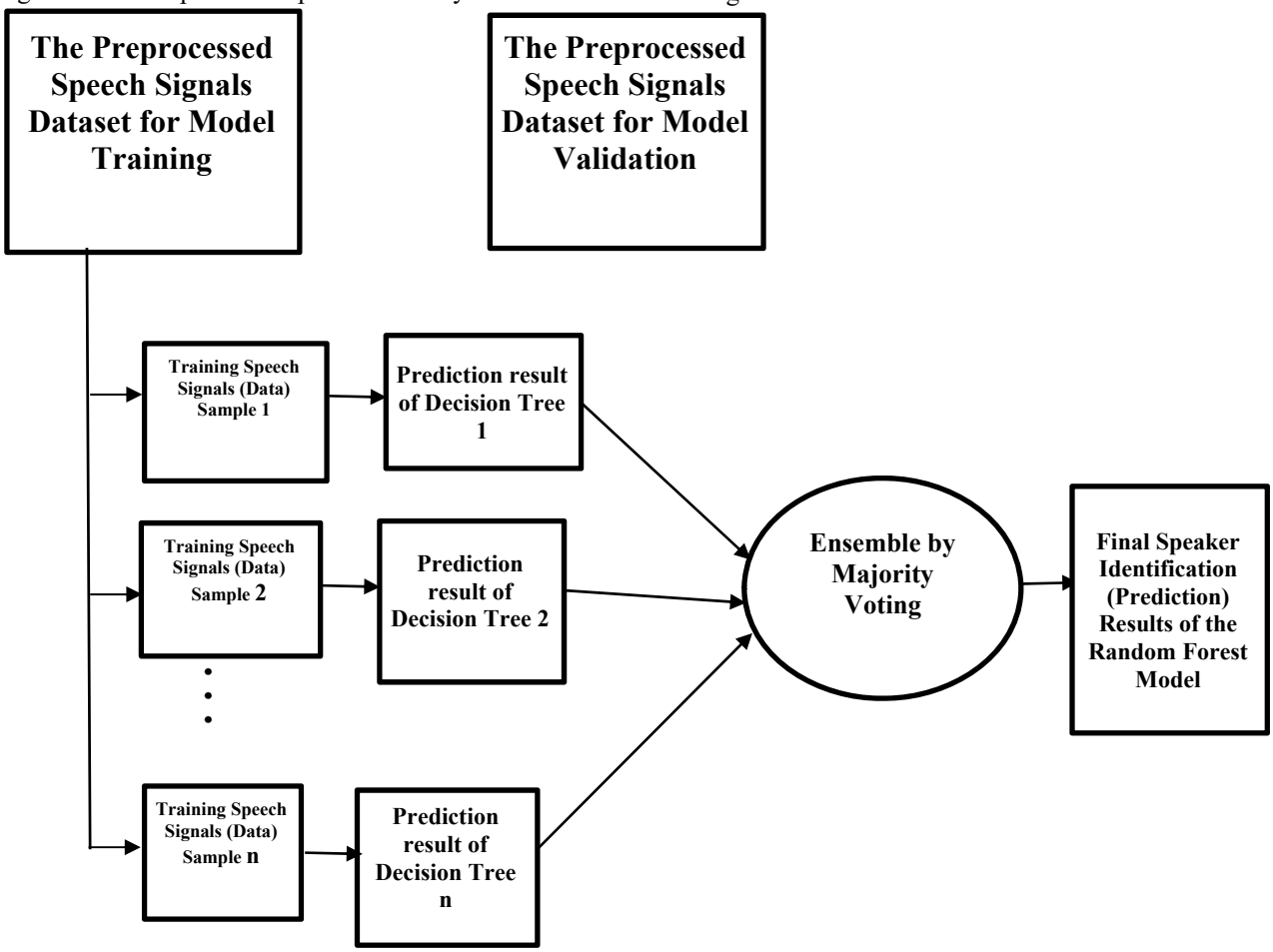

Figure 1 The architecture of the Random Forest (RF) model

In the data collection, speech samples were collected from 15 different people where the speeches were rendered in different Nigerian languages. The speech samples lasted for a maximum of 120 seconds. The acquired speech samples were properly annotated and then preprocessed to extract relevant features for the model training after which the data splitting was done and then used for the Random Forest (RF) training to identify the different speakers from the speech sample dataset. The speech samples were collected with SNR ranging from 0 dB to 30 dB while the noiseless environment with high SNR is given a finite value of 100 dB. In practice it is assumed to be infinite.

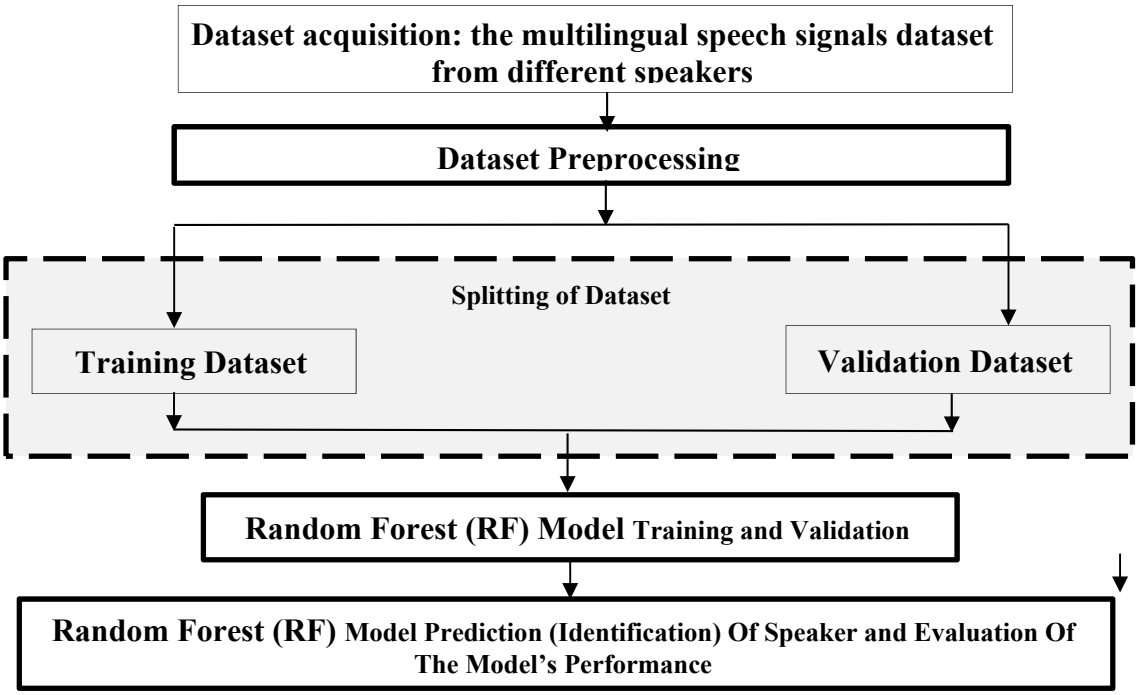

Figure 2 The flow diagram of the model training and evaluation

3. Results and discussion

In the model training and evaluation the clean speech dataset with no noise is assumed to be with SNR of 100 dB while the composite speech dataset with noise is assumed to have SNR of 10 dB. In each case the trained model is valuated using the sample speech signal with varying SNR from 0 dB to 30 dB and also with SNR of 100 dB. The result is presented for the accuracy of the model in Figure 3. The results show that when the clean signal is used for the model training it has lower accuracy when validated with the clean and composite signals whereas when the composite signal is used for the model training it has higher accuracy when validated with the clean and composite signals.

The results in Figure 3 show that the accuracy of the clean signal-trained model attained 85 % whereas the composite signal-trained model attained 96 %. Again, Figure 4 shows the percentage improvement in accuracy for using composite data trained model which showed minimum improvement of 13 % and maximum improvement of 284 %.

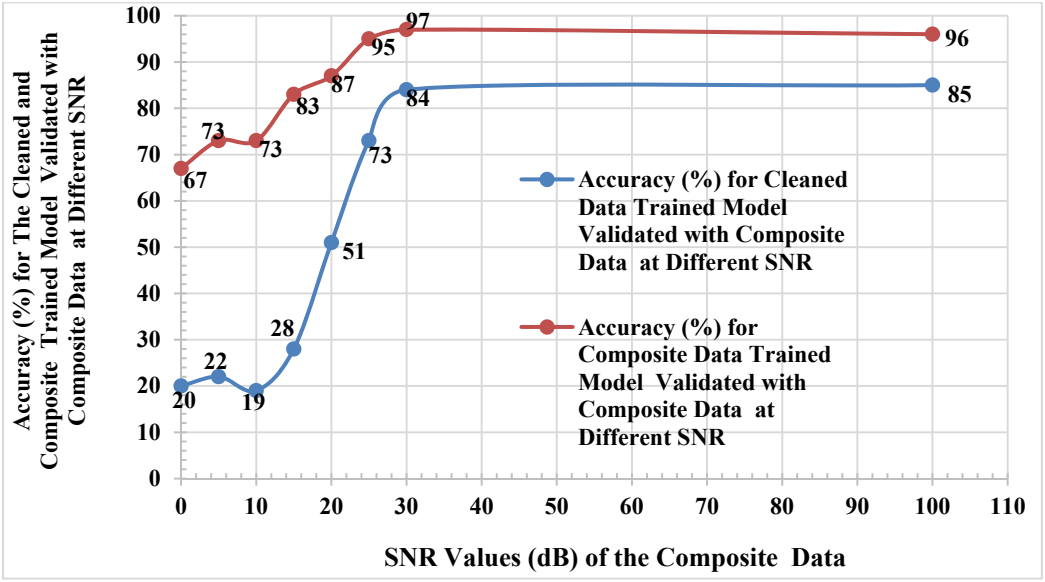

Figure 3 Scatter plot of Accuracy (%) for The Cleaned and Composite Trained Model Validated with Composite Data at Different SNR

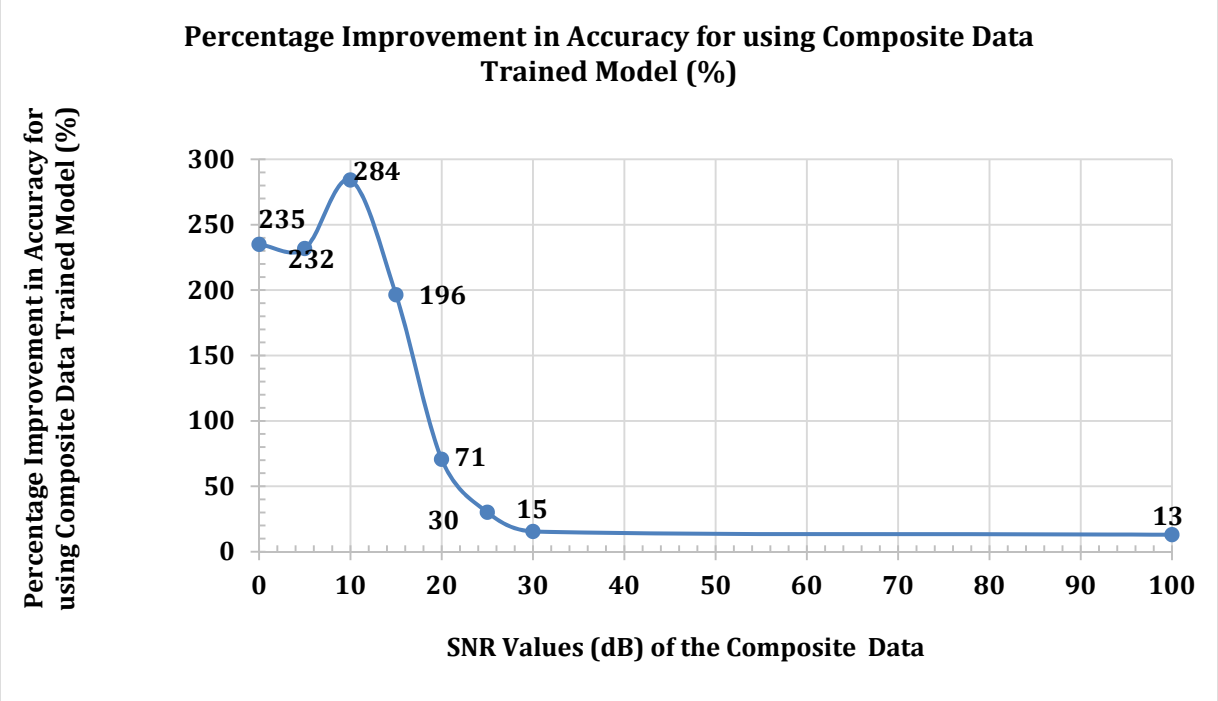

Figure 4 Scatter plot of Percentage Improvement in Accuracy for using Composite Data Trained Model (%)

The results in Figure 5 show that the precision of the clean signal-trained model attained 85 % whereas the composite signal-trained model attained 97 %. Again, Figure 6 shows the percentage improvement in precision for using composite data trained model which showed minimum improvement of 14 % and maximum improvement of 491 %.

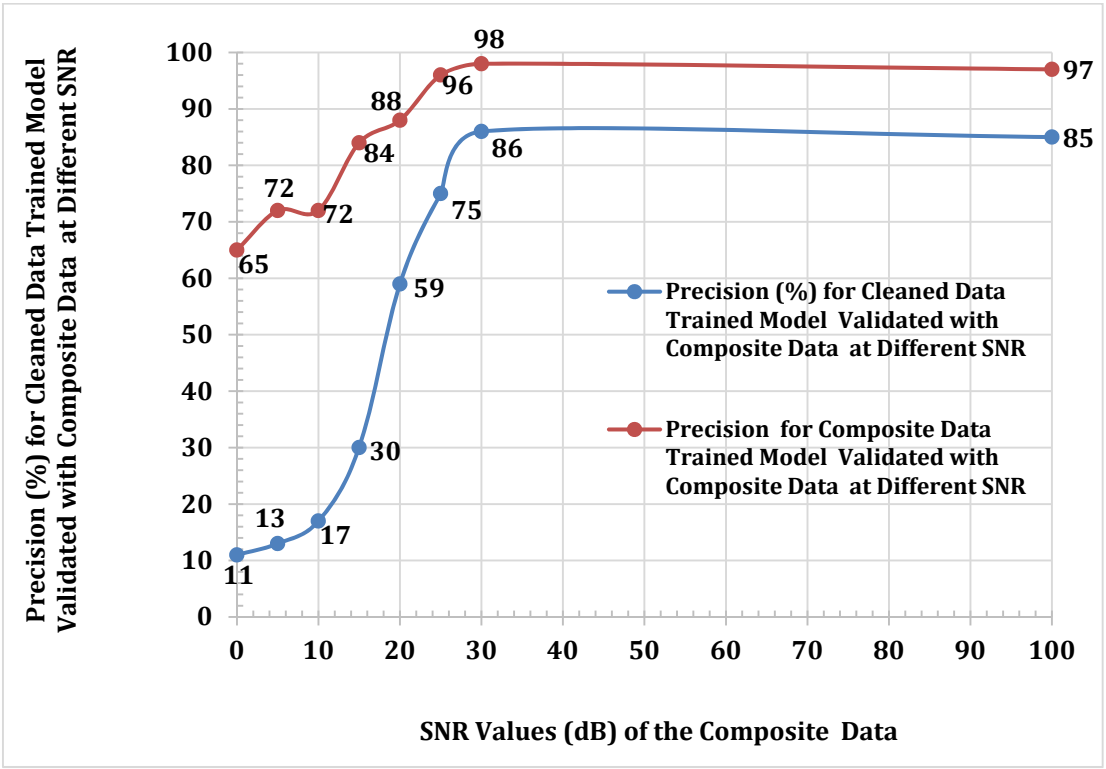

Figure 5 Scatter plot of Precision (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR

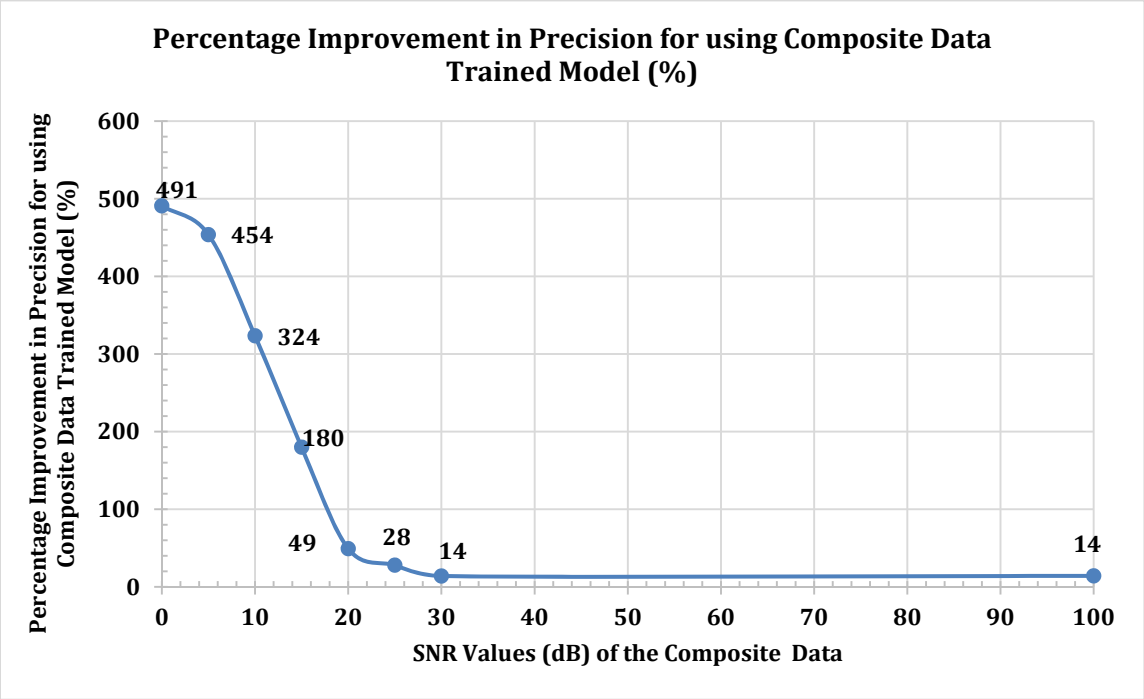

Figure 6 Scatter plot of Percentage Improvement in Precision for using Composite Data Trained Model (%)

The results in Figure 7 show that the F1_score of the clean signal-trained model attained 83 % whereas the composite signal-trained model attained 96 %. Again, Figure 8 shows the percentage improvement in F1_score for using composite data trained model which showed minimum improvement of 16 % and maximum improvement of 385 %.

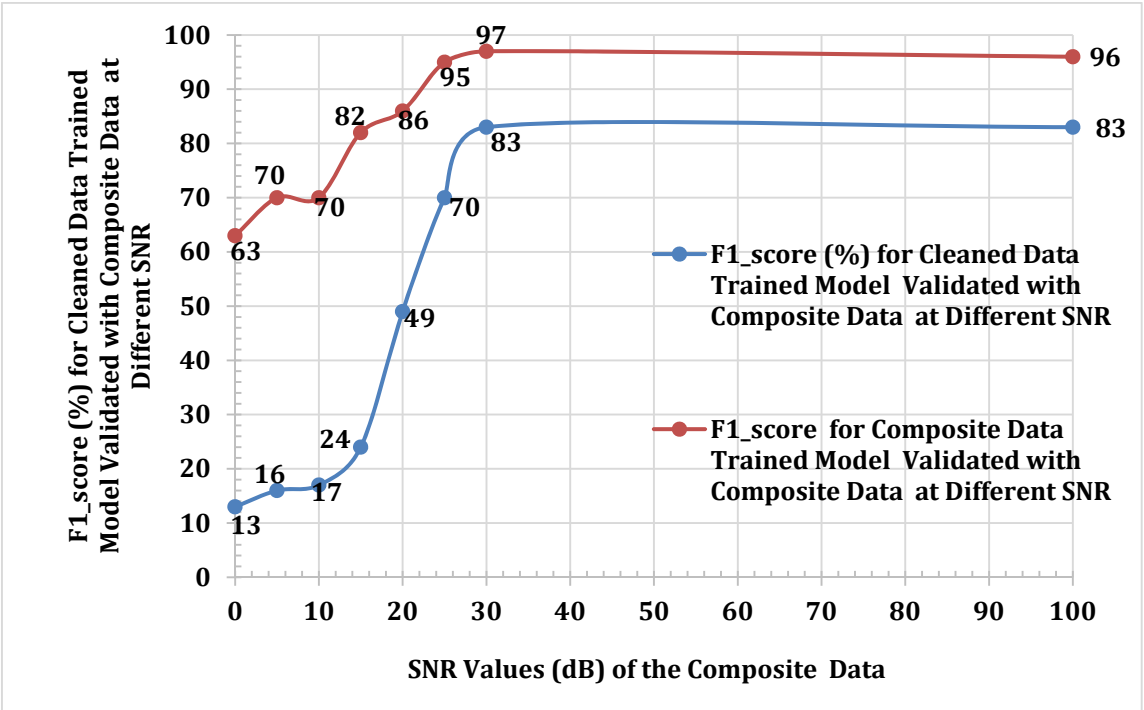

Figure 7 Scatter plot of F1_score (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR

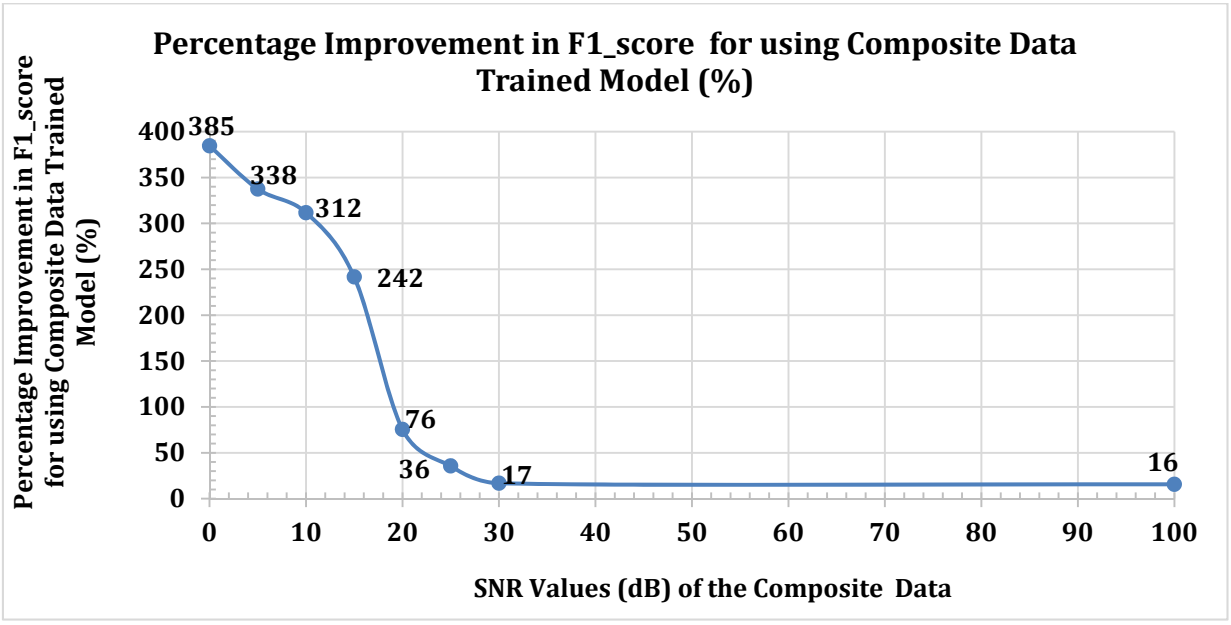

Figure 8 Scatter plot of Percentage Improvement in F1_score for using Composite Data Trained Model (%)

The results in Figure 9 show that the recall of the clean signal-trained model attained 84 % whereas the composite signal-trained model attained 96 %. Again, Figure 10 shows the percentage improvement in recall for using composite data trained model which showed minimum improvement of 14 % and maximum improvement of 284 %.

In all, training the RF model with the composite dataset ensures that model superior performance is attained in all the noise levels the test was conducted.

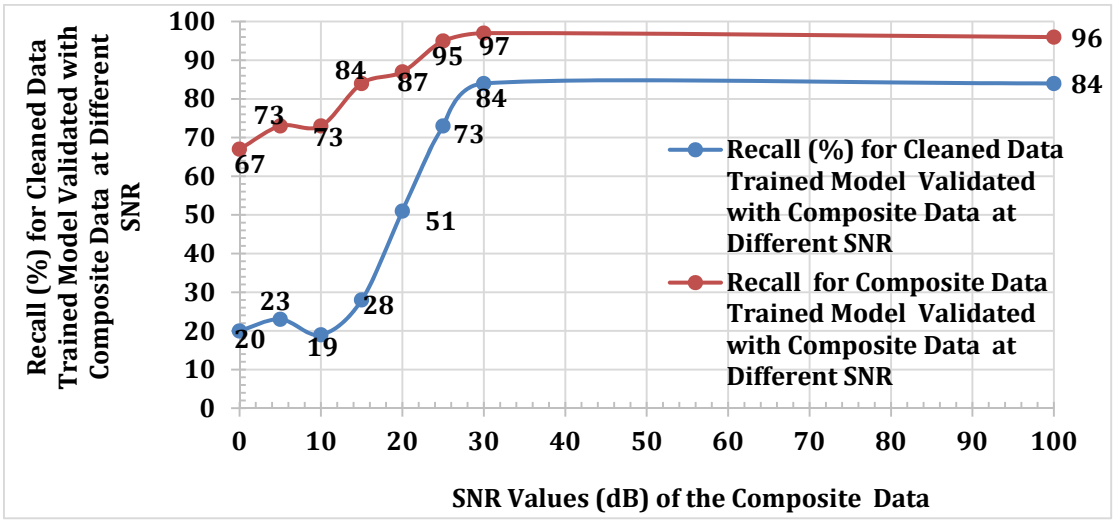

Figure 9 Scatter plot of Recall (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR

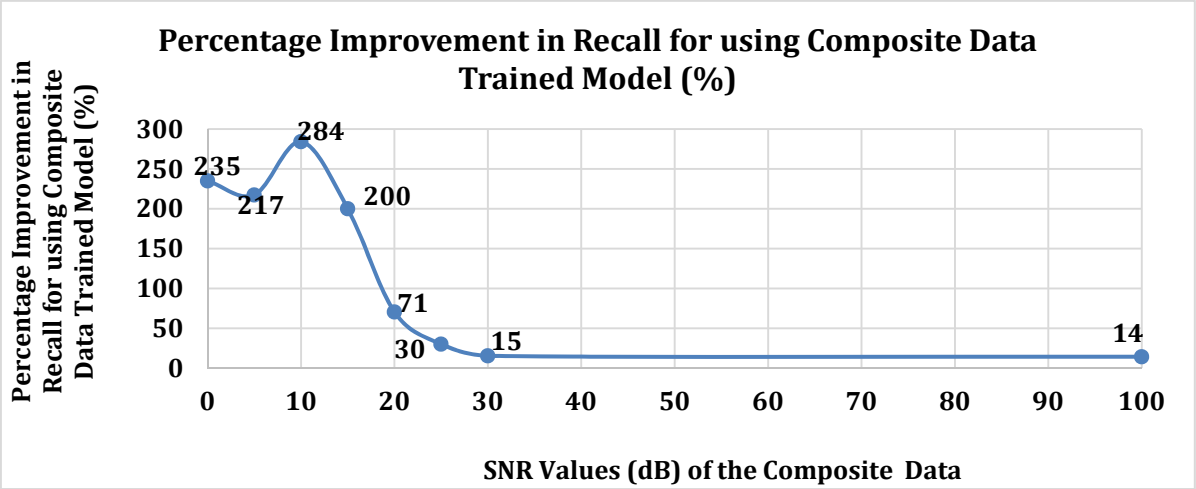

Figure 10 Scatter plot of Percentage Improvement in Recall for using Composite Data Trained Model (%)

4. Conclusion

The Random Forest model for speaker identification in a multilingual setting with different noise levels is presented. The RF model was trained with clean data and with composite dataset that has some noise. The two trained models were subjected to valuation using speaker speech signals with different noise levels and the results showed that the composite dataset-trained model ensures that the RF model has superior performance in all the noise levels the test was conducted.

References

1. Agbo, O. F. (2022). *Language Use and Code-switching among Educated English-Nigerian Pidgin Bilinguals in Nigeria* (Doctoral dissertation, Dissertation, Düsseldorf, Heinrich-Heine-Universität, 2022).
2. Okewulonu, G. G. (2023). *The Regulation of Social Media in Nigeria and its Effect on Free Speech: Perspectives from Constitutional Law and International Norms* (Doctoral dissertation, University of Saskatchewan Saskatoon).
3. Usman, K. O., Olaleye, S. B., & Clement, O. (2023). Nigerian accent-based text-to-speech program for visually impaired learners.
4. Ajimah, E. N., & Iloanusi, O. N. (2024). Biometric voice recognition system in the context of multiple languages: using traditional means of identification of individuals in Nigeria languages and English language. Res. *Stat*, *2*(1), 1-16.
5. Balasubramaniam, S., Kadry, S., Prasanth, A., & Dhanaraj, R. K. (Eds.). (2024). *AI Based Advancements in Biometrics and Its Applications*. CRC Press.
6. Defrancq, B., & Fantinuoli, C. (2021). Automatic speech recognition in the booth: Assessment of system performance, interpreters' performances and interactions in the context of numbers. *Target*, *33*(1), 73-102.
7. Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*, *14*, 1210187.
8. Zaunseder, S., Vehkaoja, A., Fleischhauer, V., & Antink, C. H. (2022). Signal-to-noise ratio is more important than sampling rate in beat-to-beat interval estimation from optical sensors. *Biomedical Signal Processing and Control*, *74*, 103538.
9. Kremer, T., Irons, T., Müller-Petke, M., & Juul Larsen, J. (2022). Review of acquisition and signal processing methods for electromagnetic noise reduction and retrieval of surface nuclear magnetic resonance parameters. *Surveys in Geophysics*, *43*(4), 999-1053.
10. Buchner, A., Hadrath, S., Burkard, R., Kolb, F. M., Ruskowski, J., Ligges, M., & Grabmaier, A. (2021). Analytical evaluation of signal-to-noise ratios for avalanche-and single-photon avalanche diodes. *Sensors*, *21*(8), 2887.
11. Mandasari, M. I., McLaren, M., & Van Leeuwen, D. A. (2012, March). The effect of noise on modern automatic speaker recognition systems. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4249-4252). IEEE.
12. Lei, Y., Burget, L., Ferrer, L., Graciarena, M., & Scheffer, N. (2012, March). Towards noise-robust speaker recognition using probabilistic linear discriminant analysis. In *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4253-4256). IEEE.
13. Matrouf, D., Kheder, W. B., Bousquet, P. M., Ajili, M., & Bonastre, J. F. (2015, August). Dealing with additive noise in speaker recognition systems based on i-vector approach. In *2015 23rd European Signal Processing Conference (EUSIPCO)* (pp. 2092-2096). IEEE.
14. Nawas, K. K., Barik, M. K., & Khan, A. N. (2021). Speaker recognition using random forest. In *ITM web of conferences* (Vol. 37, p. 01022). EDP Sciences.
15. Karthikeyan, V. (2022). Adaptive boosted random forest-support vector machine based classification scheme for speaker identification. *Applied Soft Computing*, *131*, 109826.
16. Talekar, B., & Agrawal, S. (2020). A detailed review on decision tree and random forest. *Biosci. Biotechnol. Res. Commun*, *13*(14), 245-248.