Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 12 Issue 3, March - 2025

Multilingual Speaker Identification Low Signal-To-Noise Ratio Environments Using K-Nearest Neighbour (KNN) Model

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*— Multilingual speaker identification low signal-to-noise ratio environments using k-nearest neighbour (KNN) model is presented. The study targeted identification of speaker from multilingual speech signal sampled in different Nigerian languages. The work also studied the impact of noise on the performance of the model in identifying the speaker. Several speech signals were sampled from different speakers in different Nigerian languages. Each of the speech data samples lasted for about two minutes. After detailed data preprocessing, the data were split for training and validation set and then used in the model training. Notably, the speech data were sampled under controlled noise level with signal to noise ratio (SNR) ranging from 0 dB to 30 dB and the environment with minimal noise which is referred to as the clean signal with SNR of about 100 dB. The results on the comparison of the accuracy of the composite trained model and the cleaned data trained model validated using speech data at different SNR showed that accuracy of 93 % was achieved with the composite data trained model while the clean data trained model achieved 80 % accuracy. Also, the improvement in accuracy realized by using the composite data trained model instead of the clean data trained model is about 16 % , though at some SNR the improvement was up to 880 %. Similar results were achieved in respect of precision, F1_score and recall, each showed that the composite data trained model is better. Hence, it is recommended that that the KNN model should be trained using composite dataset.**

***Keywords*— Multilingual Speaker, Signal-To-Noise Ratio, Speaker Identification, K-Nearest Neighbour (KNN) Model, Classification Models**

1. Introduction

Today, multilingual speaker identification system has become a very useful tool for many applications [1,2,3]. Also, as globalization increases language diversity in communication, multilingual systems play an essential role in overcoming the complexities of cross-lingual and multi-accented speech [4,5]. However, over the years, one of the most significant challenges facing multilingual speaker identification lies in handling linguistic diversity, including differences in phonemes, accents, intonation, and speech patterns across various languages [6,7]. This requires the system to be adaptable and robust, ensuring accurate identification regardless of the speaker's language or dialect [8,9]. By incorporating advanced voice analysis techniques and machine learning models, these systems can deliver reliable and consistent results in diverse and dynamic environments [10,11].

Multilingual speaker identification systems are designed to determine the identity of an unknown speaker by analyzing their voice and comparing it against a database of known speakers, even when multiple languages are involved [12,13]. Unlike speaker verification, which involves confirming or rejecting a claimed identity through a one-to-one comparison, speaker identification answers the broader question, "Who is speaking?" This process involves key stages such as extracting unique features from the speaker's voice, matching these features to patterns, and

comparing them to stored voice samples [14]. These systems are particularly valuable in applications where identifying the speaker is critical, such as law enforcement, forensic analysis, and customer service operations.

While multilingual system has been studied for many languages, the Nigerian languages have not really been used because of lack of appropriate dataset for such study. Moreover, the study of speech identification system in noisy environment is essential since the impact of noise power can be very significant in the model performance [16,17]. Hence, in this study the K-Nearest Neighbour (KNN) model is used for speaker identification in noisy environment with different signal to noise ratio [18,19]. The study targeted the Nigerian languages. The focus is to evaluate the performance of different versions of the model under different noise levels**.**

2. Methodology

The motivation in this work is to use K-Nearest Neighbours (KNN) for identification of speaker from multilingual speech signal sampled in different Nigerian languages. The work also studied the impact of noise on the performance of the model in identifying the speaker.

Notably, the K-Nearest Neighbours (KNN) is a non-parametric classifier that classifies a new sample based on the majority class of its K-nearest neighbours in the feature space. The distance between samples is typically measured using Euclidean distance $d(x_i, x_j)$ between two points x_i and x_j given as:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - x_{j,k})^2} \quad (1)$$

The class label is determined by the most frequent label among the k nearest points. The details of the procedure used in the KNN model is given in Algorithm 1. The research procedure for the KNN model training and performance evaluation is presented in Figure 1.

Algorithm 1: The KNN Procedure

Step 1: Input relevant data items

Step 1.1: Input the training dataset and the test data item

Step 1.2: Input K // the number for the nearest neighbours

Step-2: Compute the Euclidean distance, $d(x_i, x_j)$ between the test data item and all the training dataset.

Step-3: Arrange the Euclidean distance between the test data and all the training dataset in ascending order (from the smallest to the largest distance).

Step-4: Take the first K neighbours in the sorted list (they are the K nearest neighbours based on the Euclidean distance

Step-5: From the selected K neighbours, count the number of data points that occurred for in each of the data categories.

Step-6: By voting method, the category of the test data point is the category with the highest count in Step 5

Step-7: Repeat the Step 3 to Step 5 for all the Test Data Items.

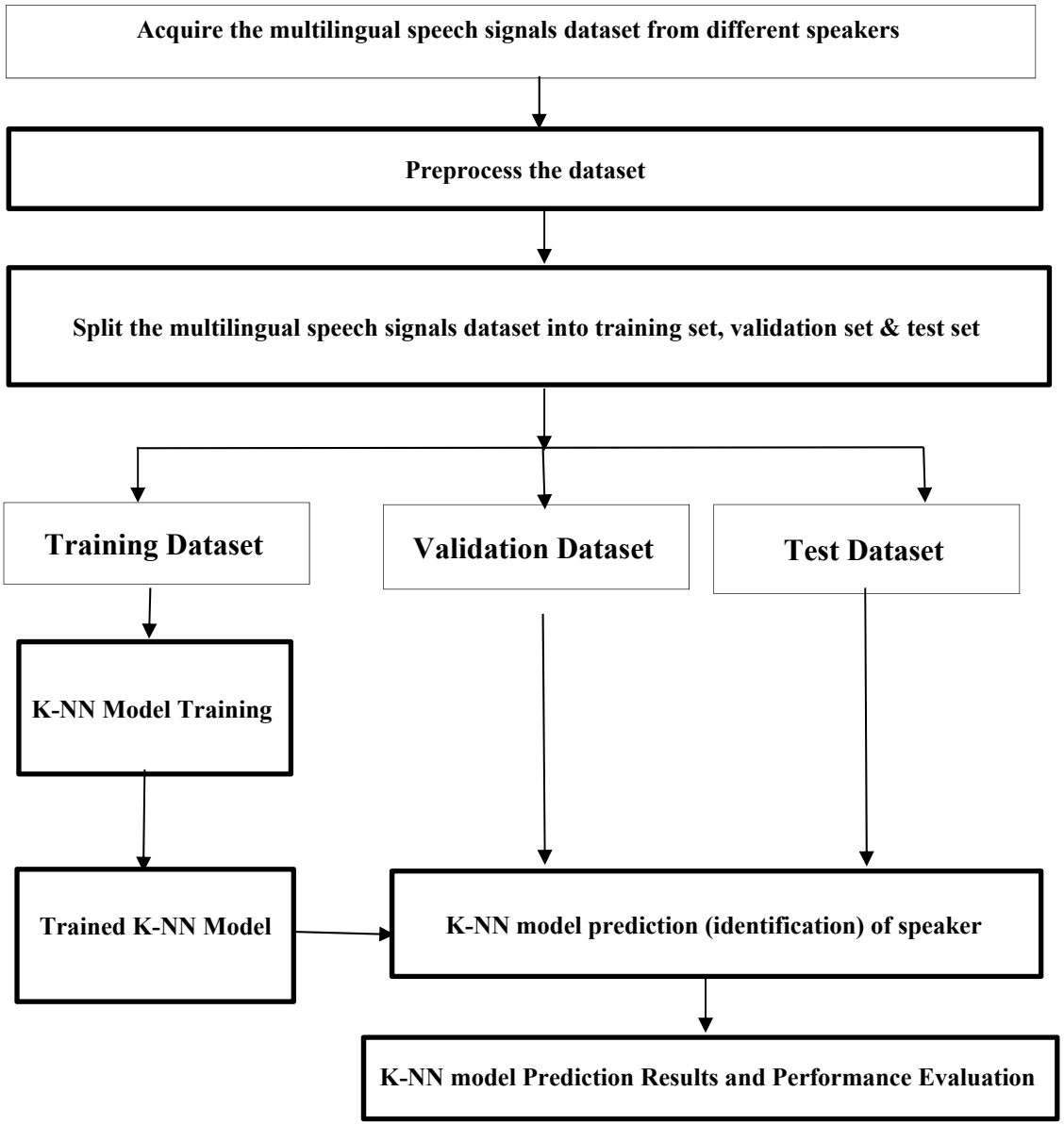


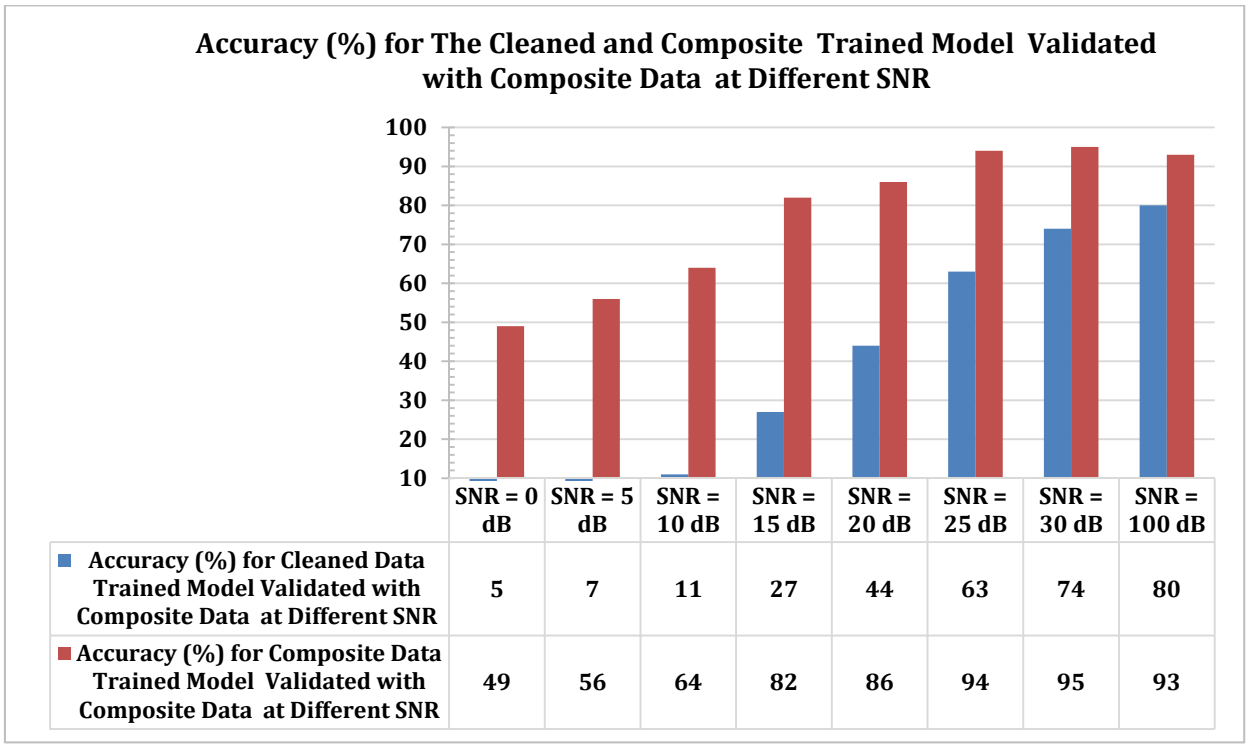
Figure 1 The research procedure for the KNN model training and performance evaluation

In order to accomplish the main aim of this study, speech signals are sampled from different speakers speaking in different Nigerian languages. Each of the speech data samples lasted for about two minutes. Detailed data preprocessing was done which included data augmentation, feature extraction, data normalization, and then data splitting into the training and validation datasets. The speech data were sampled under controlled noise level with signal to noise ratio (SNR) ranging from 0 dB to 30 dB and the environment with minimal noise which is referred to as the clean signal with SNR of about 100 dB. The speech sample with low SNR are referred to as composite signal with high presence of noise.

4.1.3 Performance Evaluation of k-Nearest Neighbour (KNN) Model

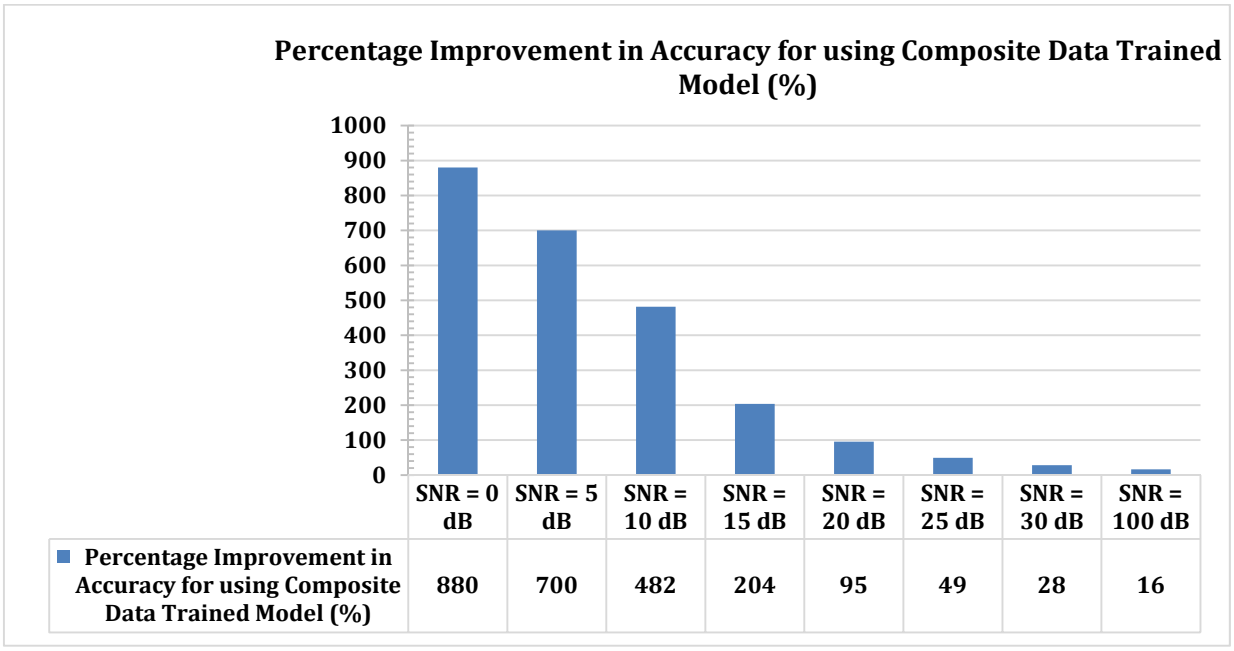
The results detailing the performance of the k-Nearest Neighbour (KNN) model under varying conditions are presented in Figure 2 to Figure 8. The bar chart comparing the accuracy (%) of the composite trained model and the cleaned data trained model validated using speech data at different SNR is shown in Figure 2. The results in Figure 2 showed that accuracy of 93 % was achieved with the composite data trained model while the clean data trained model achieved 80 % accuracy. In Figure 3 the improvement in accuracy realized by using the composite data trained model instead of the clean data trained model is about 16 % , though at some SNR the improvement was up to 880 %.

3. Results and discussion



	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Accuracy (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR	5	7	11	27	44	63	74	80
Accuracy (%) for Composite Data Trained Model Validated with Composite Data at Different SNR	49	56	64	82	86	94	95	93

Figure 2 The bar chart comparing the accuracy (%) of the composite trained model and the cleaned data trained model validated using speech data at different SNR

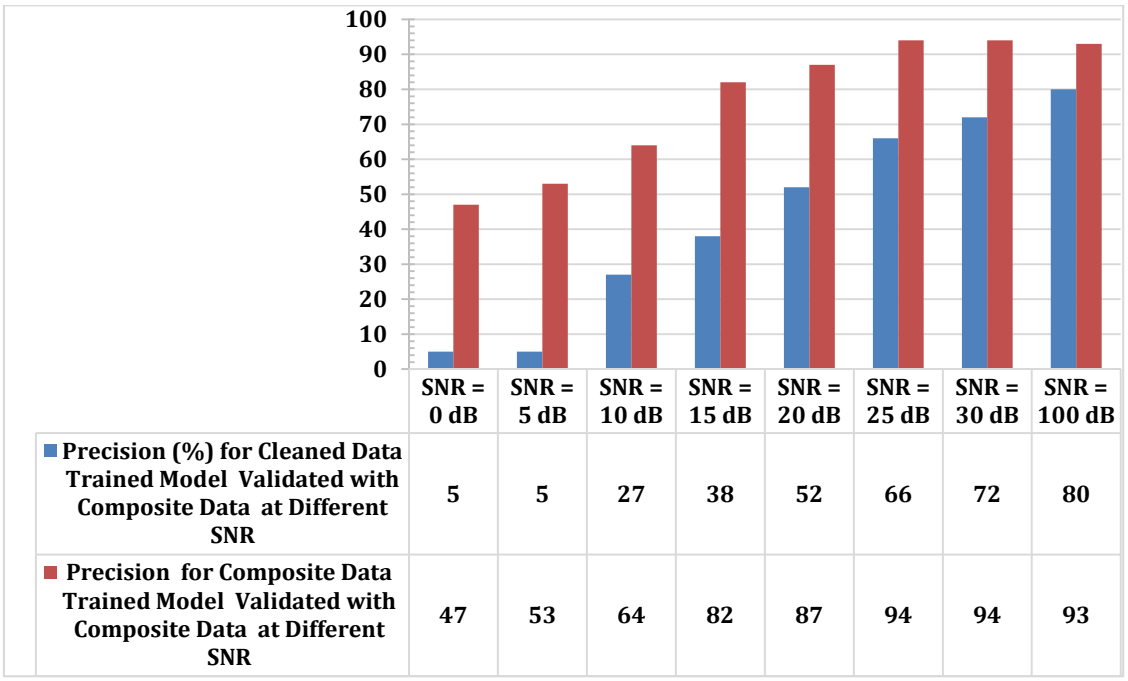


	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Percentage Improvement in Accuracy for using Composite Data Trained Model (%)	880	700	482	204	95	49	28	16

Figure 3 The bar chart summarizing the improvement in Accuracy realized by using the composite data trained model instead of the clean data trained model

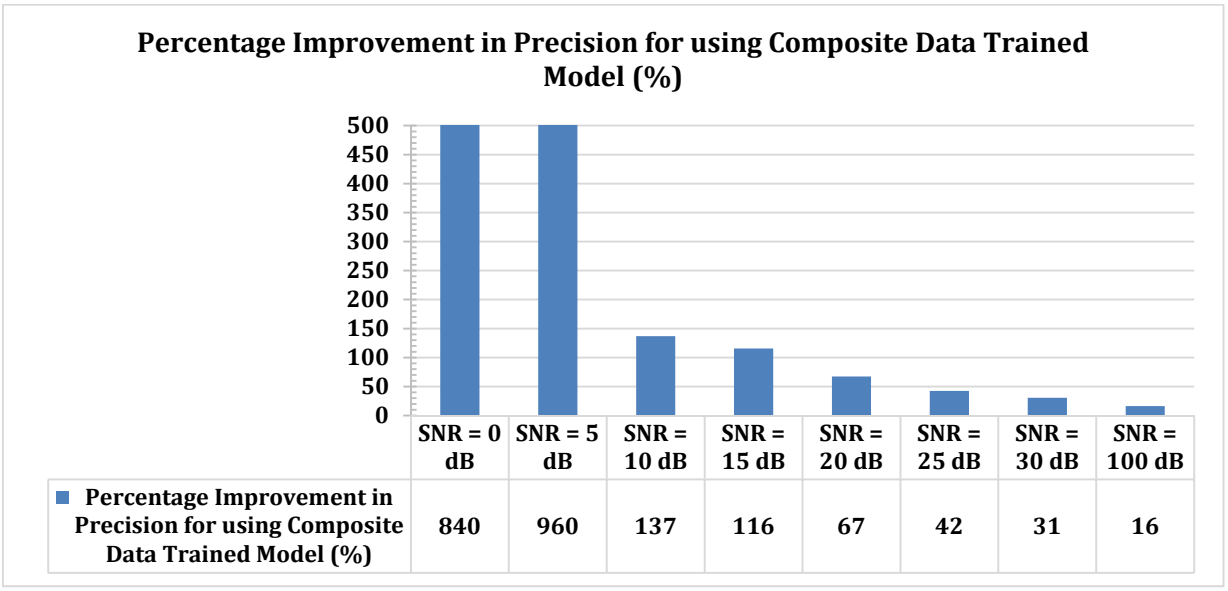
The bar chart comparing the precision of the composite trained model and the cleaned data trained model validated using speech data at different SNR is shown in Figure 4. The results in Figure 4 showed that precision of 93 % was achieved with the composite data trained model while the clean data trained model achieved 80 % precision. In Figure 5 the improvement in precision realized by using the composite data trained model instead of the clean data trained model is about 16 % , though at some SNR the improvement was up to 960 %.

The bar chart comparing the F1_Score of the composite trained model and the cleaned data trained model validated using speech data at different SNR is shown in Figure 6. The results in Figure 6 showed that F1_Score of 92 % was achieved with the composite data trained model while the clean data trained model achieved 79 % F1_Score. In Figure 7 the improvement in F1_Score realized by using the composite data trained model instead of the clean data trained model is about 16 % , though at some SNR the improvement was up to 1000 %.



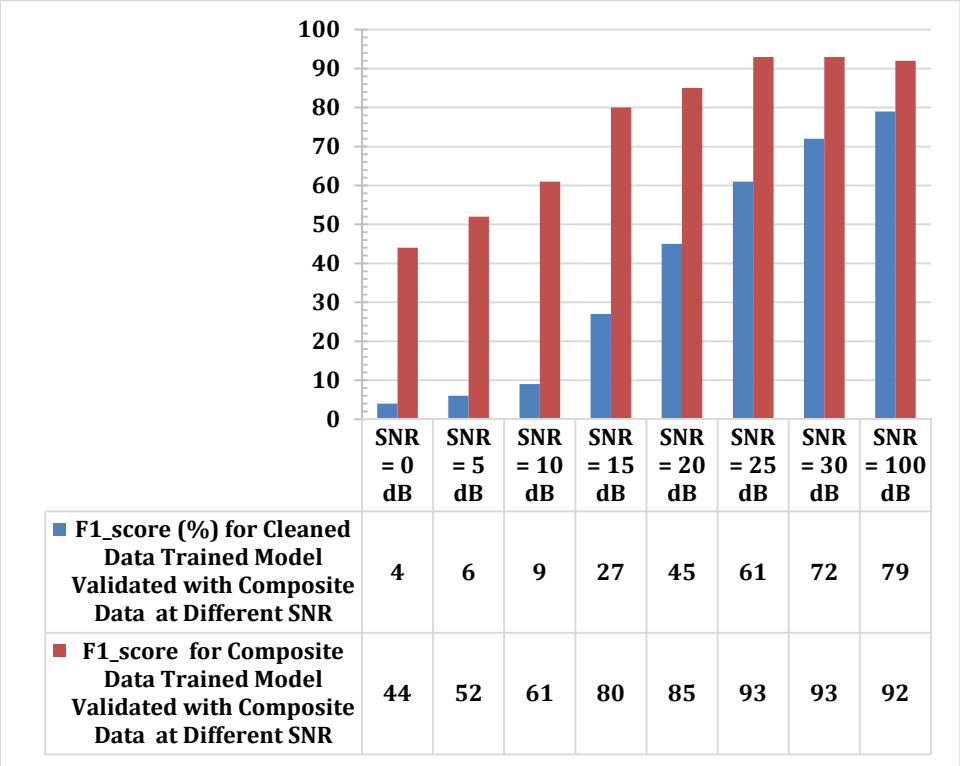
	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Precision (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR	5	5	27	38	52	66	72	80
Precision for Composite Data Trained Model Validated with Composite Data at Different SNR	47	53	64	82	87	94	94	93

Figure 4 The bar chart comparing the Precision (%) of the composite trained model and the cleaned data trained model validated using speech data at different SNR



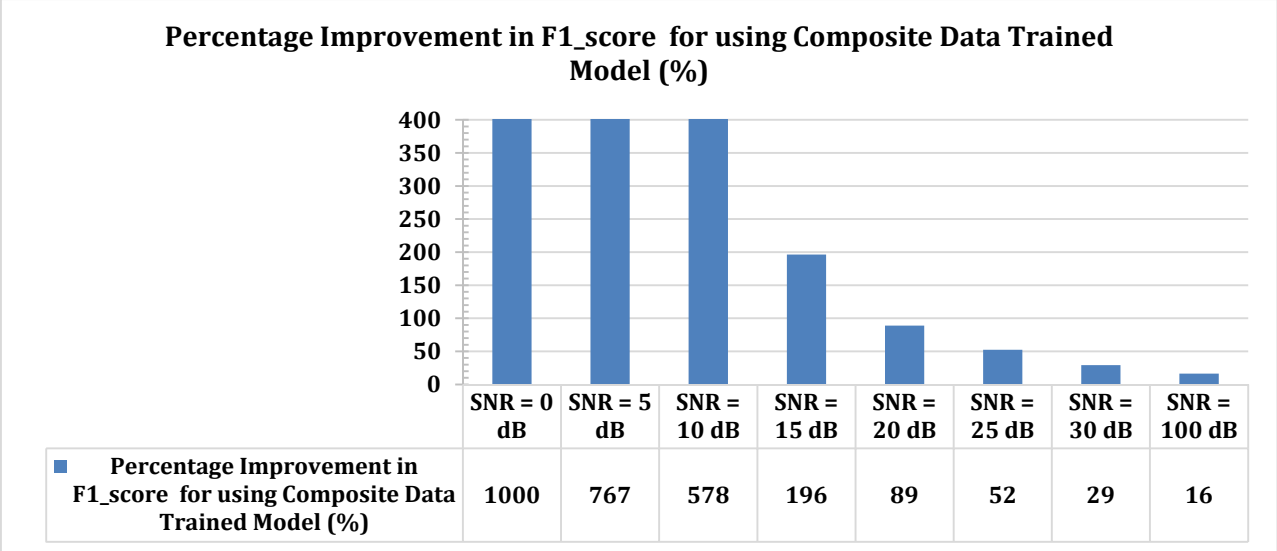
	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Percentage Improvement in Precision for using Composite Data Trained Model (%)	840	960	137	116	67	42	31	16

Figure 5 The bar chart summarizing the improvement in Precision realized by using the composite data trained model instead of the clean data trained model



	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
F1_score (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR	4	6	9	27	45	61	72	79
F1_score for Composite Data Trained Model Validated with Composite Data at Different SNR	44	52	61	80	85	93	93	92

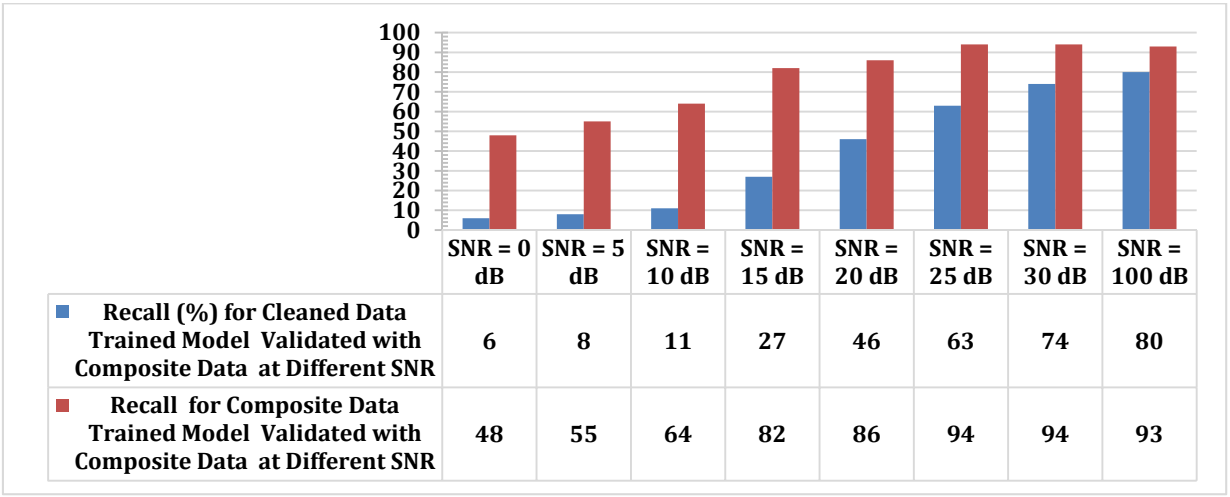
Figure 6 The bar chart comparing the F1_Score (%) of the composite trained model and the cleaned data trained model validated using speech data at different SNR



	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Percentage Improvement in F1_score for using Composite Data Trained Model (%)	1000	767	578	196	89	52	29	16

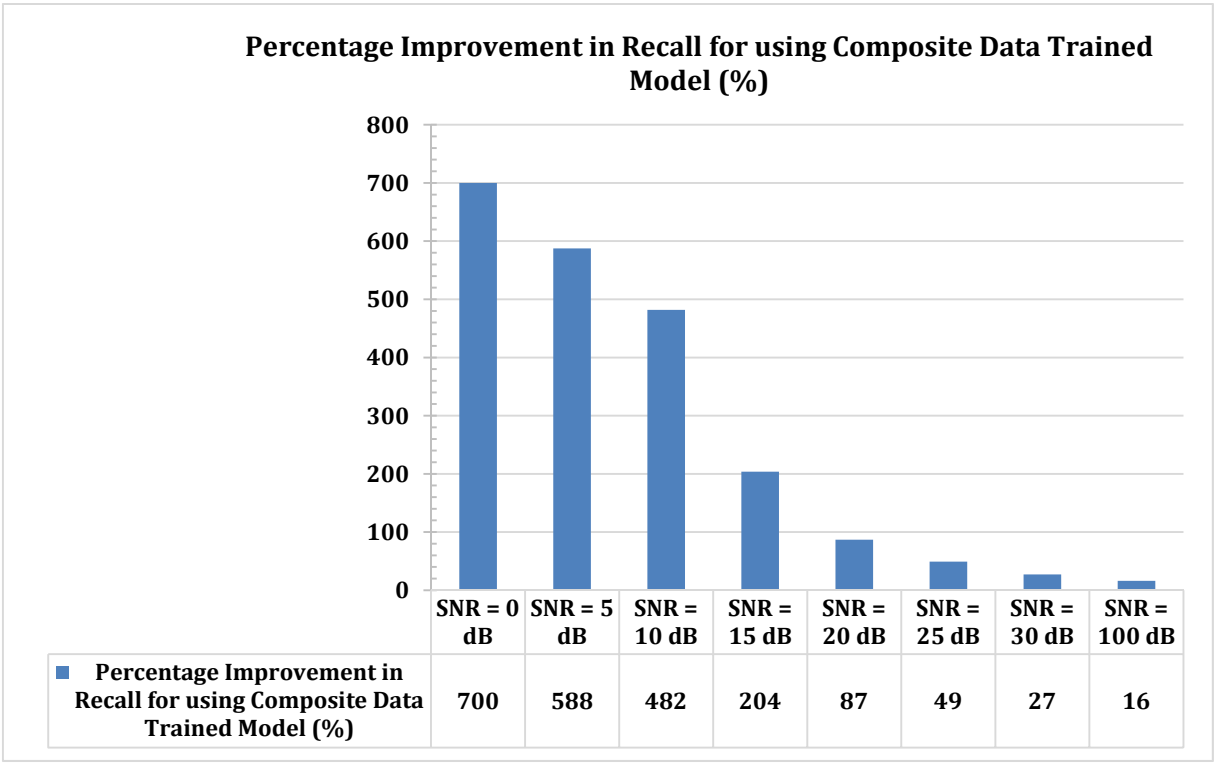
Figure 7 The bar chart summarizing the improvement in F1_score realized by using the composite data trained model instead of the clean data trained model

The bar chart comparing the Recall of the composite trained model and the cleaned data trained model validated using speech data at different SNR is shown in Figure 8. The results in Figure 8 showed that Recall of 93 % was achieved with the composite data trained model while the clean data trained model achieved 80 % Recall. In Figure 9 the improvement in Recall realized by using the composite data trained model instead of the clean data trained model is about 16 % , though at some SNR the improvement was up to 700 %.



	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Recall (%) for Cleaned Data Trained Model Validated with Composite Data at Different SNR	6	8	11	27	46	63	74	80
Recall for Composite Data Trained Model Validated with Composite Data at Different SNR	48	55	64	82	86	94	94	93

Figure 8 The bar chart comparing the Recall (%) of the composite trained model and the cleaned data trained model validated using speech data at different SNR



	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB	SNR = 15 dB	SNR = 20 dB	SNR = 25 dB	SNR = 30 dB	SNR = 100 dB
Percentage Improvement in Recall for using Composite Data Trained Model (%)	700	588	482	204	87	49	27	16

Figure 9 The bar chart summarizing the improvement in Recall realized by using the composite data trained model instead of the clean data trained model

4. Conclusion

The K-Nearest Neighbours (KNN) model is presented for recognizing speaker from a dataset of multilingual speech signal presented for different Nigerian languages. The study examined the model training in the presence of different noise power levels depicted using signal to noise ratio (SNR). The results showed that the model that was trained using the composite speech signal with significant noise power performed better that the model that was trained using clean signal with negligible noise power. In essence, the KNN is recommended to be trained with composite speech signal to enhance the prediction performance of the model.

References

1. Waibel, A., Geutner, P., Tomokiyo, L. M., Schultz, T., & Woszczyna, M. (2002). Multilinguality in speech and spoken language systems. *Proceedings of the IEEE*, *88*(8), 1297-1313.
2. Schultz, T., & Kirchhoff, K. (Eds.). (2006). *Multilingual speech processing*. Elsevier.
3. Stein-Smith, K. (2016). The role of multilingualism in effectively addressing global issues: the sustainable development goals

and beyond. *Theory and Practice in Language Studies*, *6*(12), 2254-2259.

4. Crespo, D. G. O., Holgado, M. C., & Nanquil, L. (2021, December). The Impact of Multilingualism on Global Education and Language Learning: A Book Review. In *Linguistic Forum-A Journal of Linguistics* (Vol. 3, No. 4, pp. 3-4).
5. Ushioda, E. (2017). The impact of global English on motivation to learn other languages: Toward an ideal multilingual self. *The Modern Language Journal*, *101*(3), 469-482.
6. McLeod, S., Verdon, S., Baker, E., Ball, M. J., Ballard, E., David, A. B., ... & Zharkova, N. (2017). Tutorial: Speech assessment for multilingual children who do not speak the same language (s) as the speech-language pathologist. *American Journal of Speech-Language Pathology, 26(3)*, 691-708.
7. Alshehri, A., & AlShabeb, A. (2023). Exploring attitudes, identity, and linguistic variation among Arabic speakers: Insights from acoustic landscapes. *International Journal of Arabic-English Studies*, 24(2), 1-16.
8. Togneri, R., & Pullella, D. (2011). An overview of speaker identification: Accuracy and robustness issues. *IEEE circuits and systems magazine*, 11(2), 23-61.
9. Togneri, R., & Pullella, D. (2011). An overview of speaker identification: Accuracy and robustness issues. *IEEE circuits and systems magazine*, *11*(2), 23-61.
10. Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN computer science*, *2*(6), 1-20.
11. Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of computational methods in engineering*, *27*, 1071-1092.
12. Tirumala, S. S., Shahamiri, S. R., Garhwal, A. S., & Wang, R. (2017). Speaker identification features extraction methods: A systematic review. *Expert Systems with Applications, 90*, 250-271.
13. Mittal, A., & Dua, M. (2022). Automatic speaker verification systems and spoof detection techniques: review and analysis. *International Journal of Speech Technology, 25(1)*, 105-134.
14. Turner, H. (2021). *Security and privacy in speaker recognition systems* (Doctoral dissertation, University of Oxford).
15. Li, J., Deng, L., Gong, Y., & Haeb-Umbach, R. (2014). An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(4), 745-777.
16. Zhang, Z., Geiger, J., Pohjalainen, J., Mousa, A. E. D., Jin, W., & Schuller, B. (2018). Deep learning for environmentally robust speech recognition: An overview of recent developments. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *9*(5), 1-28.
17. Kacur, J., Vargic, R., & Mulinka, P. (2011, June). Speaker identification by K-nearest neighbors: Application of PCA and LDA prior to KNN. In *2011 18th International Conference on Systems, Signals and Image Processing* (pp. 1-4). IEEE.
18. Safi, M. E., & Abbas, E. I. (2023). Speech recognition algorithm in a noisy environment based on power normalized cepstral coefficient and modified weighted-KNN. *Eng. Technol. J*, *41*, 1107-1117.
19. Safi, M. E., & Abbas, E. I. (2023). Speech recognition algorithm in a noisy environment based on power normalized cepstral coefficient and modified weighted-KNN. *Eng. Technol. J*, *41*, 1107-1117.