ARID ZONE JOURNAL OF ENGINEERING, TECHNOLOGY &

Published by the Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria.
Print ISSN: 1596-2644, Electronic ISSN: 2545-5818 www.azojete.com.ng

ENVIRONMENT

ORIGINAL RESEARCH ARTICLE

EXPLORING THE INFLUENCE OF NOISE ON VOICE RECOGNITION SYSTEMS: A CASE STUDY OF SUPERVISED LEARNING ALGORITHMS

M. O. Balogun*, B. Jimada-Ojuolape and M. O. Mahmoud

[1]Department of Electrical and Computer Engineering, Kwara State University, Malete, Ilorin, Kwara State, Nigeria.*
Department of Electrical and Computer Engineering, Kwara State University, Malete, Ilorin, Kwara State, Nigeria.
**Corresponding author's email address: Monsurat.balogun@kwasu.edu.ng*

ABSTRACT:
Speech recognition systems have become increasingly prevalent in various applications, ranging from virtual assistants to voice-controlled devices and dictation software. The desire to design human-computer interfaces that are more accessible, intuitive, and efficient cannot be overemphasized. Such advancements hold significant potential to enhance accessibility, productivity, safety, and user experience across various domains and industries.. Despite significant advancements in speech recognition technology, several challenges persist. One primary challenge is achieving high accuracy and robustness across diverse speaking styles, accents, and environmental conditions. Background noise, variations in pronunciation, and overlapping speech can introduce errors and degrade performance, especially in real-world scenarios. This study explores the pursuit of accurate and lightweight algorithms in biometric applications, focusing on voice recognition. Using a dataset featuring renowned leaders' voices, the research compares K-Nearest Neighbors (KNN) and Artificial Neural Network (ANN) techniques. Employing Mel Frequency Cepstral Coefficients (MFCC) and considering noisy and noiseless environments, the study reveals that ANN outperforms KNN. In noisy conditions, ANN achieves 62.4% accuracy, while KNN reaches 33.4%. In noiseless settings, ANN's accuracy rises to 95%, surpassing KNN's 88%. Assessment metrics like False Acceptance Rate, False Rejection Rate, F1 Score, Recall, Precision, and Receiver Operator Characteristics Curve are analyzed. The study emphasizes the detrimental impact of noise on recognition accuracy and underscores ANN's consistent superiority over KNN. Despite challenges posed by noisy environments, the research highlights the potential benefits of these approaches, emphasizing ANN's superior performance across scenarios. This research showcases the significance of accurate biometric systems, emphasizing ANN's advantages in enhancing usability and precision, especially in real-world noisy conditions.

Keywords: Accuracy, Identification, Speech Recognition, Pre-processing, Voice Biometric

1. Introduction

Speech serves as a primary mode of human communication, facilitating the expression of thoughts and emotions. The fundamental process of speech production involves the interplay of the lungs, vocal folds (voice box), and articulators (Tanberk and Tükel, 2022). The collaboration between vocal folds and articulators generates a diverse range of intricate sounds, with the potential to convey varying emotional tones, enhancing effective

communication (Alattar *et al*., 2023). While every individual's voice is distinct due to factors like body size, shape, and vocal cord characteristics, voice patterns are not entirely unique (Gupta *et al.,* 2023). During verbal interactions, the vibrations of human vocal cords convey critical information about voice identity, enabling people to appropriately respond to speakers by avoiding overly warm greetings to strangers or overly cold responses to friends (Al-Zakarya and Al-Irhaim, 2023).

Conventionally, humans recognize each other's speech by developing mental models for different speakers based on their exposure to diverse voices (Dhole *et al.,* 2023). These models incorporate fundamental differences in features such as rhythm and tone. By naturally adapting and familiarizing themselves with various speaker models, individuals learn to identify and comprehend distinct speakers. While voice/speech recognition is a crucial aspect of human identification, conventional voice recognition methods may fall short, particularly in security contexts (Jawalkar *et al.,* 2023). Traditional techniques are susceptible to mimicry, wherein an individual imitates another's voice to gain unauthorized access, raising significant security concerns, especially for applications demanding robust user authentication (Booysens and Viriri, 2022). Li *et al.,* (2023) says that traditional voice recognition encounters limitations including variability in speech patterns, environmental noise, adverse conditions like illness or fatigue, limited adaptability, cross-linguistic challenges, and more. In response, researchers and developers have been focusing on advanced methods like deep learning, neural networks, and refined data collection techniques to establish more robust and versatile voice recognition systems, such as voice recognition biometric systems which can surmount the challenges (Kao & Chueh, 2023).

Voice biometrics, also known as speaker recognition, is a technology that focuses on identifying or verifying individuals by analyzing their unique vocal characteristics. (Minaee *et al.,* 2023). This technology encompasses two primary modes: speaker identification and speaker verification. Speaker identification involves matching a person's voice against a voiceprint database to ascertain their identity. In contrast, speaker verification entails comparing a provided voiceprint with the voiceprint linked to the claimed identity for verification purposes (Ammour *et al.,* 2023). The latter mode is commonly used for security applications, access control, and phone-based authentication.

Voice biometric systems hold a pivotal role in human identity authentication and access control within security setups (Alharbi and Alshanbari, 2023). Their history dates to the early 1950s with the creation of "AUDREY." The first documented speech recognizer was developed by Davis, Biddulph, and Balashek at Bell Labs, capable of recognizing speakers by their pronunciation of numbers from 1 to 9 (Wendehorst *et al.,* 2021). Contemporary voice recognition systems like "SIRI" by Apple and "Alexa" by Amazon have integrated voice recognition into consumer platforms, demonstrating a perception of intelligence and contextual language understanding (Dvořák *et al.,* 2021). The inherent variations in human voices are valuable traits in biometric security systems, and as technology advances, the demand for more intelligent and efficient biometric systems intensifies (Anil and Ravikumar, 2022).

M. K. and Aithal, (2022) said that efficient functioning of voice biometric systems involves extracting key characteristics from a speaker's voice, preprocessing these characteristics, and then distinguishing the speaker based on these attributes in comparison to other enrolled speakers. This process generally entails employing statistical algorithms on preprocessed voice data (Benmachiche *et al.,* 2022). Selecting an appropriate statistical model algorithm for a voice recognition system considers factors like hardware, design specifications, and implementation costs. Utilizing simpler yet high-performing algorithms can significantly reduce setup costs, false match rates, and false non-match rates.

The emergence of machine learning algorithms like artificial neural networks (ANN) and artificial immune systems (ANS) has introduced a biological approach to pattern recognition (Xue and Zhou, 2023). Usually, in audio data machine learning applications, traditional machine learning algorithms such as Support Vector Machines (SVC), Logistic Regression, and K Nearest

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

Neighbours tend to have pre-processed data with signal domain features such as the Time domain features like the amplitude enveloping, zero crossing rate, and the root mean square energy (Hitelman *et al.,* 2022) . ANN, for example, mirrors human brain neurons with nodes organized into layers, interconnected through nodes that culminate in the output or decision layer. Node activation hinges on a mathematical function involving input node activations from the preceding layer, weight values, and biases. Techniques can enhance these parameters to minimize the cost function and enhance performance (Pudurkar *et al.,* 2022). Although ANNs possess advantages like automatic feature extraction, adaptability with incomplete data, and recognition of intricate relationships, their drawbacks include long training times, hardware dependency, and the complexity of interpreting what the network learns (Can and Atilgan, 2023)

In contrast, K-Nearest Neighbors (KNN) is a widely used classification algorithm known as a "lazy learner" as it relies on training vector data points for decision-making without learning the relationships between them. KNN computes Euclidean distances for a predefined K-number of training vectors categorized with specific classes (Imdad *et al.,* 2017). The computed distances identify the closest neighbors to a test point vector, and the decision about the test point's class is determined by the majority rule among the nearest neighbor's data points (Demidova, 2021). While KNN offers simplicity, easy implementation, and flexibility for multi-modal classes, it relies on manual feature development compared to the automatic feature evolution in ANN (Murthy *et al.,* 2015)

Biometric identification systems strive to meet the demand for more efficient and lightweight algorithms in voice/speech recognition. Improving the accuracy of these systems is crucial for enhancing user experience on platforms employing biometric security measures. Achieving this goal entails minimizing False Match Rates and False Non-Match Rates while simultaneously increasing accuracy and deploying lightweight algorithms for swift user identification. To address these objectives, this study conducts a comparative analysis between two prominent algorithms: K Nearest Neighbors (KNN) and Artificial Neural Networks (ANN). The aim is to identify the most suitable model capable of meeting the evolving demands in speech recognition. Both KNN and ANN serve as classifiers in this study, and their performances are evaluated using a comprehensive set of evaluation metrics. These metrics include Accuracy, False Acceptance Rate (FAR), False Rejection Rate, Confusion Matrix, Recall, Precision, Area Under the Curve (AUC), F1 Score, and Receiver Operating Characteristic (ROC) analysis.

Much research has been carried out on how to improve performance learning algorithms. When it comes to classification in biometric systems, choosing the best on the best classifier has always been an issue for the past decades. To improve the performance of biometric systems, many algorithms have been developed and modified by researchers, among which were reviewed in this section.

Conventional Automatic Speech Recognition (ASR) models typically consist of three fundamental components: an acoustic model, a pronunciation dictionary, and a language model (Xu *et al.,* 2020). The Acoustic Model (AM) assesses the likelihood of acoustic units, such as phones or sub-word units, which can be represented using Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs). (Juang *et al.,* 1991) Typically, Gaussian Mixture Models (GMMs) are employed to estimate the probability distribution of phones within individual states, while Hidden Markov Models (HMMs) are utilized to determine transition probabilities between states. Each state corresponds to an acoustic event, such as a phone. The GMM-HMM model undergoes training via the expectation maximization (EM) technique, and Viterbi decoding is applied to identify the optimal state sequence within HMMs. In natural utterances, the pronunciation of phones often varies based on acoustic context. Hence, context-dependent triphone HMMs are utilized to model speech sounds, wherein each phone is represented with both left and right contexts. In recent advancements in Automatic Speech Recognition (ASR) models, Gaussian Mixture Models (GMMs) have been supplanted by deep neural networks (DNNs). (Hinton *et al.,* 2012).

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

Hybrid DNN-HMM models are commonly used in ASR, combining components like the Language Model (LM) to compute word sequence probabilities and enhance acoustic model accuracy. LMs leverage linguistic knowledge from text corpora, implicitly learning syntactic and semantic rules to refine acoustic model hypotheses. A pronunciation dictionary aligns phonetic transcriptions with raw text for LM use. These components are independently trained and combined using Finite State Transducers (FSTs) to construct a search graph. Decoders generate lattices, scored, and ranked to produce target word sequences. ASR model evaluation involves metrics like word error rate (WER) and phone error rate (PER) to gauge performance.

Modern ASR Modern ASR systems are fully end-to-end; for example, (Amodei *et al.,* 2016) This architecture employs an encoder–decoder framework, where the input audio undergoes processing through a series of convolutional layers, resulting in the creation of a condensed vector representation. Subsequently, the decoder utilizes this encoded vector as input to generate a sequence of characters. Several different objective functions such as CTC (Graves *et al.,* 2006), ASG (Collobert *et al.,* 2016.), Transduction Differentiable decoding (Collobert *et al.,* 2019) can be used to optimize end-to-end ASR. Moreover, different architectures of neural networks such as ResNet (Wang *et al.,* 2016.), TDS (Hannun *et al.,* 2019) and Transformer (Shaw *et al.,* 2018) have been explored. The output labels of the end-to-end system can be characters or sub-word units such as byte-pair encoding (BPE). An external LM can be incorporated to improve the overall system performance. One of the challenges in ASR, even if supervised, is the variability that is characteristic of natural spoken utterances due to speaker and environmental conditions. Utterances by different speakers have acoustic differences that can be difficult to disentangle from the phonetic content. Even for an individual speaker, variability arises due to speaking rate, intensity, affect, and so on.

Furthermore, ASR models are often trained on clean speech data, but they are often evaluated on real-time noisy speech data. Sources of noise include background noises and signal distortions through the input device. Speaker-independent models could be trained using data that includes multiple speakers, but this often degrades the performance of ASR and requires larger amounts of data for training to achieve decent performance. The same applies to background noises and different environmental conditions. Instead, state-of-the-art ASR systems are speaker-adaptive: they capture the variability of speakers using IVectors (Saon *et al.,* 2013) and X-Vectors (Snyder, 2020) which are low-dimensional vectors that encode speaker-specific features. In addition, various augmentation techniques can be utilized to supplement the training data with more examples that reflect the expected variability in test conditions. For example, volume and speed perturbation are used to capture the variability between utterances. Similarly, noise-augmentation is used to supplement the training data with different environmental conditions (Ko *et al.,* 2017). All these strategies are combined to train robust multi-condition ASR systems that can handle multiple sources of variability in speech.

Kaensar, (2013) developed a hand digit recognition by comparing the performance of three machine learning algorithms: The Neural Network (ANN), Support Vector Classifier (SVC), and K nearest neighbors. The dataset used was the Optical Recognition of Handwritten Digits Data set from UCI machine learning repository, which consists of a total number of 5,620 handwritten digits. The results produced showed that SVM was the best classifier to recognize handwritten digits with 96.93% recognition rates, however SVM has slow training time compared with ANN and SVM.

Murthy and Meenakshi, (2015) compared the classification of Artificial Neural Network (ANN), Support Vector Machine (SVM) and K Nearest Neighbor (KNN) on cardiac ischemia based on morphological features that were extracted from ECG signals. The experiment was conducted by finding the optimal number of hidden neurons in the Feed forward ANN as well as using the most accurate kernel for the SVM varying the K value that would give the best performance for the K Nearest Neighbors. The experimental results showed that ANN outperformed SVM and KNN with a testing classification accuracy of 96.62%. It was

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

recommended that the morphological features can still be optimally selected to improve performance on the Artificial Neural Network

Lanjewar *et al.,* (2015) performed a comparison of the accuracy of Gaussian Mixture Models (GMM) and K-Nearest Neighbor (KNN) on speech emotion recognition systems. The authors utilized the spectral components of Mel Frequency Cepstral Coefficients (MFCC), wavelet features of speech, and the pitch of vocal traces based off voices that fall under six emotional categories such as happy, angry neutral, surprised, fearful, and sad, using the database of the Standard Berlin emotion database (BES). The authors concluded that the Gaussian Mixture Models technique produced the better accuracy by recognizing the angry category with the highest rate of 92%, it had a general low performance for the "surprise" category with 25% accuracy. KNN had its overall strength in classifying the happy emotion category with 90% accuracy, it has poor performance on both fear and surprise categories with 34% and 25% accuracy respectively. The paper then used the F measure and precision evaluation metrics to present an overall model performance comparison of the two models with Gaussian Mixture models being the better overall model in speech emotion recognition but recommends that to build a more efficient and robust model these two algorithms may be hybridized to balance their strengths and weaknesses.

Richardson *et al.,* (2015) developed a Deep Neural Network Approaches to Speaker and Language Recognition": This study delves into the application of deep neural networks for both speaker and language recognition. The authors explore deep architecture such as Deep Belief Networks (DBN) and Long Short-Term Memory (LSTM) networks. The research showcases the potential of deep learning in learning complex relationships within voice data, contributing to more accurate speaker recognition.

In Richardson *et al.,* (2015) ASVspoof: The Automatic Speaker Verification Spoofing and Countermeasures Challenge. The paper describes the ASV Spoofing and Countermeasures (ASVspoof) initiative, which aims to fill this void. Through the provision of a common dataset, protocols, and metrics, ASVspoof promotes a sound research methodology and fosters technological progress. This paper also describes the ASVspoof 2015 dataset, evaluation, and results with detailed analyses. A review of post-evaluation studies conducted using the same dataset illustrates the rapid progress stemming from ASVspoof and outlines the need for further investigation. The authors emphasize the significance of anti-spoofing measures and the integration of multiple features for robust speaker recognition systems.

Minaee *et al.,* (2023), developed a Voice Biometric Recognition Using Deep Learning Techniques. This study explores the efficacy of deep learning techniques, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), for voice biometric recognition. The researchers demonstrate that deep learning models exhibit improved performance in feature extraction, leading to enhanced accuracy in speaker identification. The study underscores the potential of neural networks in tackling the challenges of variability in speech patterns and background noise.

He *et al.,* (2020) worked on an End-to-End Text-Independent Speaker Verification with Triplet Loss. This research investigates the application of triplet loss in end-to-end text-independent speaker verification. The authors propose a framework that learns discriminative embedding directly from raw audio signals. The study emphasizes the importance of proper loss functions in training deep networks for voice biometrics. The results highlight the potential of this approach for reducing false acceptance rates and improving system security.

Nigam *et al.,* (2022) worked on Biometric Authentication for Intelligent and Privacy-Preserving Healthcare Systems. This research investigates the applicability of voice biometrics in healthcare and security domains. The authors highlight the potential of voice biometrics for patient identification, access control, and telemedicine applications. The study underscores the non-intrusive nature of voice-based authentication and its contribution to efficient and secure healthcare systems.

Moosavian *et al.,* (2013) carried surveys on a new scheme for fault diagnosis of main journal-bearings of internal combustion (IC) engine based on power spectral density (PSD) technique

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

and two classifiers, namely, KNN) and ANN. Vibration signals for three different conditions of journal-bearing; normal, with oil starvation condition and extreme wear fault were acquired from an IC engine. PSD was applied to process vibration signals. Thirty features were extracted from the PSD values of signals as a feature source for fault diagnosis. KNN and ANN were trained by training data set and then used as diagnostic classifiers. Variable K value and hidden neuron count (N) were used in the range of 1 to 20, with a step size of 1 for KNN and ANN to gain the best classification results. The roles of PSD, KNN and ANN techniques were studied. The best performance was 94.5% and 90.5% on training and testing data set which belonged to the N = 5 for ANN with PSD.

The reviewed works collectively highlight the growing interest and advancements in biometric voice recognition. Deep learning techniques, including neural networks, are prominently featured, showcasing their ability to address challenges related to variability in speech patterns and noise. Hence, this work compared the performance of ANN and KNN on noiseless and noisy voice dataset.

2 . Materials and Methods

The performance of ANN and KNN on voice recognition was evaluated in this work. The Kaggle dataset was used, the dataset consists of segmented speech voice data from five prominent speakers namely: Nelson Mandela, Benjamin Netanyahu, Jens Stoltenberg, Julia Gillard, and Margaret Thatcher (*Kaggle Dataset for Speech Recognition - Google Search*, n.d.). The dataset consists of 1500 audio clip samples, each with a second duration and having 16000-sample rate PCM encoded. The dataset also consists of background noise samples to increase the difficulty of the speaker recognition tasks. These noise samples consist of audience laughing and clapping, dish washing sounds, running tap and a person imitating a cat's sound while collecting the speech from the speakers. The acquired audio data were pre-processed into a suitable data format that is suitable for the algorithms to train and test. The developed biometric system was simulated using Python library called Librosa. The Mel Frequency Cepstral Coefficients (MFCC) are extracted from the pre-processed data. KNN and ANN were used as classifiers. At classification stage, the algorithms are subjected to a query voice with an enrolled speaker and identity of a speaker is based on the decision of the algorithms.

The first step is the audio feature extraction of the voice signal from the dataset. Extraction of compatible and most salient features of a trait improves the overall performance of a classification algorithm (Balogun *et al.,* 2022). The K nearest neighbour is initially trained on the time domain features and the frequency domain features, this is because features independent is an important fact that helps in discriminating against audio. The Artificial Neural Network trained only Mel-Frequency Cepstral Coefficients (MFCCs), these MFCC is known to convey the different values that construct the formants of audio, which in this work is the human voice. ANN can produce the timbre of a sound with its time and frequency domain feature which captures all possible signal domain relationships that would be very suitable for deep learning models. However, Mel-Frequency Cepstral Coefficients (MFCCs) is used to optimize the performance of the two algorithms. Shown in Figures 1 and 2 is the flowchart and system flow diagram of the developed biometric voice recognition system.

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

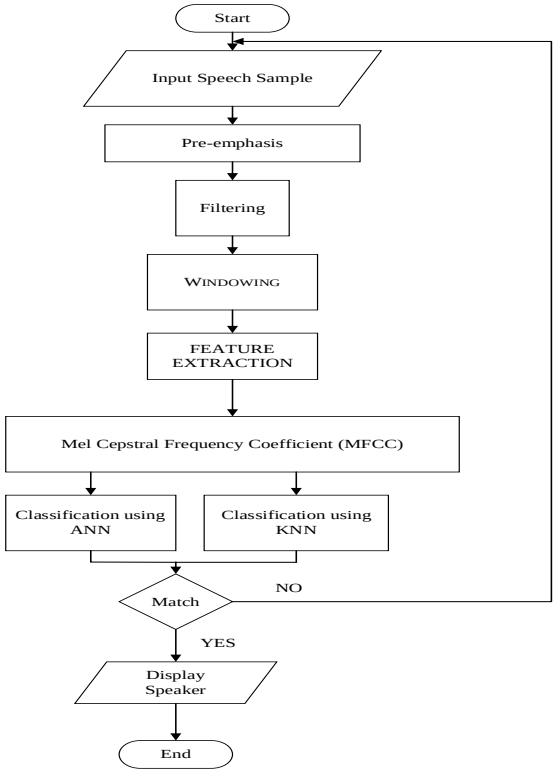


Figure 1: Flow Diagram of the Developed Biometric Voice Recognition System

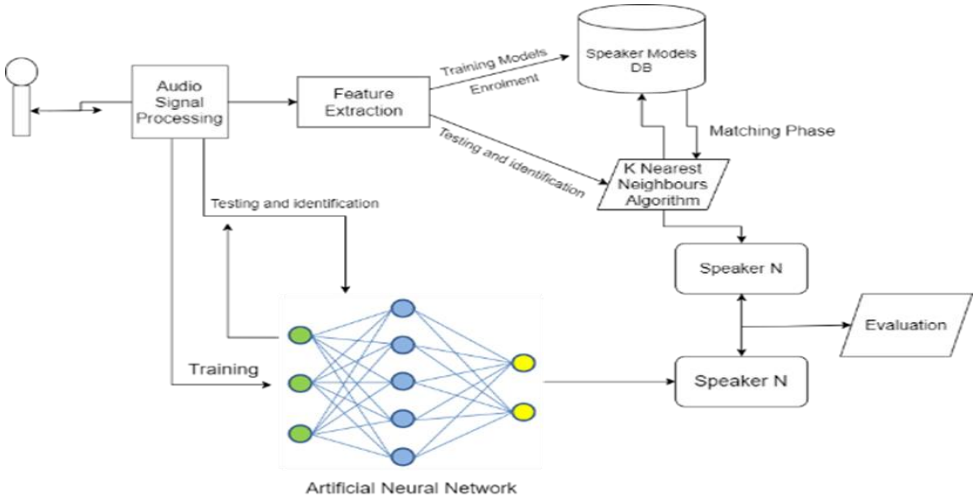


Figure 2: System Flow Diagram of the Developed Voice Recognition System

3.1. Stages of the Developed Speech Recognition System

The system consists of Several units including:

- Speech sample: presents the data previously acquired into the whole system shown as a microphone.
- Audio Signal Processing unit: performs necessary noise reductions and reduces the speech sample into useful features that the algorithms can easily be trained and tested on.
- Feature extraction unit: extraction of the salient features that would be needed for the K nearest neighbour algorithm was carried out at this stage.
- Speaker Model database: the enrolment of speakers and provision of training data set for the K nearest neighbours occurred at this stage.
- Classification Stage: KNN and ANN algorithms are used as classifiers in this work.

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

Evaluation unit: the performance of the two algorithms on voice recognition was evaluated at this stage. ANN was implemented using a python library called KERAS, due to the variation in the number of hidden layers that can be implemented, a Deep Neural Network (DNN) was considered for the work. With Stochastic Gradient Descent (SGD) used as the optimizer which depends on the performance of the model when tried with either, Sigmoid used as the activation function for the neural network and Mean Squared Error (MSE) used as the cost function. The pseudocode to demonstrate the implementation of ANN as follows:

3.2. Pseudocode for feature classification of ANN

1: BEGIN data preparation:
2: Let X be an array of processed Training Voice data of size $(m * n)$ consisting of $(x_i,\ y_i)$ data points where x_i are data points and y_i is the label for corresponding point;
3: Let Y be the array labels for records in $X;\ Y \in (1, 2, 3\ \ldots . c)$ where c is the number of speakers contained in the dataset;
4: INITIALIZE neural network values by assigning randomized variables within a distribution in a small range;
5: ACTIVATE the back propagation to update the network correctly by passing the input data x_1, x_2, x_3, x_n with its respective output labels y_1, y_2, y_3, y_n
6: CALCULATE the actual outputs of the neurons in the hidden layer using the Sigmoid activation function.
Where output of a neuron being $y_i(p) = sigmoid\left(\left[?_{j=1}^{n} x_i(p) * w_{ij} - \mathrm{T}_j\right]\right)$
7: CALCULATE the actual outputs of neurons at the output layer that decides which speaker based on data passed in;
8: CALCULATE the error gradient for the neurons at the output layer;
9: CALCULATE the weight corrections using Mean Squared Error as cost function, gradient descent and backpropagation with momentum and learning rate values to compute update values;
10: UPDATE the weights at the hidden neurons with corrected values;
11: ITERATE from Step 5 until accuracy is enhanced and overfitting is becoming apparent;
12: END.

The K nearest neighbor algorithm was initiated from a python library called scikit learn through the object sklearn.neighbors.KNeighborsClassifier, where the necessary parameters were set like the n-neighbors which sets the k-value subset of data and metric which sets the distance metric used to compute the nearest neighbors. Heuristics like cross validation was used to determine the optimal k-value and metric to be used. The pseudocode to demonstrate the implementation of KNN is as follows:

3.3 Pseudocode for feature classification of KNN

1: // Begin data preparation:
2: Let X be an array of processed training voice data of size $(m * n)$ consisting of $(x_i,\ c_i)$ data points where x_i are data points and c_i is the label for corresponding point.;
3: Let C be the array labels for records in $X;\ C \in (1, 2, 3\ \ldots . c)$ where c is the number of speakers in dataset;
4: Let x_j be an unknown test point to find class c_j;
5: // Begin Algorithm Execution:
6: CALCULATE $\boldsymbol{d}(x_j, x_i)$, where i= 1, 2, 3 ,.. n are datapoints, and $\boldsymbol{d}$ is the Euclidean distance metric between all known points x_i and the unknown test point x_j;
7: SORT a list (D) containing the calculated distances $\boldsymbol{d}(x_j, x_i)$, in an ascending order;
8: Let k be a positive integer to take the first k number of distances in D;
9: Let K be a list of k values of D;

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

10: Let KNN be a list that stores the corresponding $x_i$points corresponding to values in K list;
11: Let $K_n$be the number of points belonging to a class n within the array KNN where $n = (1,2,3, \dots c)$ where c is the number of classes in C;
12: Let RANK store all the K_n values present in KNN array;
13: FOR value in RANK
14: IF value in RANK greater than all values
15: OUTPUT unknown variable x_j to be in class value (K_n) in RANK
16: END.
17: END.
18: END.

3.4 Performance Evaluation Metrics

The performance of the algorithms was compared using True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), False Acceptance Rate (FAR). False Rejection Rate (FRR), Accuracy, Precision, recall, F1_Score and Receiver Operating Characteristic Curve. True Positive (TP) is an experimental outcome where the algorithm correctly predicts the positive class. True Negative (TN) is an experimental outcome where the model correctly predicts the negative class. False Positive (FP) is an experimental error in binary classification in which a classifier incorrectly recognized a class, while False Negative (FN) is like false positive but in binary classification it incorrectly indicated the absence of a class.
False Acceptance Rate (FAR) is likelihood that the system will incorrectly identify an impostor as a legitimate speaker. İt is the ratio of false positives to the number of identification attempts.

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad [1]$$

False Rejection Rate measures the likelihood that the system will fail to identify a legitimate speaker. A lower FRR indicates a higher level of accuracy because it means that the system is less likely to mistakenly identify individuals. (FRR) is a ratio of false negative to the identification attempts.

$$FRR = \frac{FN}{TP + FN} \quad [2]$$

Accuracy is the general term used to describe the overall performance of a classification method. The formula for getting accuracy in relation to other confusion metrics is;

$$\text{Accuracy } (\%) = \left({}^{(\text{TP} + \text{TN})}/_{(\text{TP} + \text{FN} + \text{TN} + \text{FP})}\right) * 100\% \quad [3]$$

Precision is a metric that captures the ability of a model to correctly label a sample as true positive. In speaker recognition, precision measures the times the model was able to correctly identify a speaker by not misclassifying an impostor as the right speaker.

$$\text{Precision } (\%) = \left[\left({}^{\text{TP}}/_{\text{TP} + \text{FP}}\right)\right] * 100\% \quad [4]$$

Recall measures the ability of a system to correctly identify all relevant instances, typically within a specific class or category. İt is the ratio of the number of true positives to the number of false negatives. It measures the classifiers capacity to locate all the positive samples.

$$\text{Recall } (\%) = \left[\left({}^{\text{TP}}/_{\text{TP} + \text{FN}}\right)\right] * 100\% \quad [5]$$

F1_Score measures the harmonic mean of the precision and recall, it is a more suitable metric in measuring the accuracy of a model, since it considers imbalanced data and uses the average of the precision and recall rates.

$$F1Score = 2 * {}^{Precision * recall}/_{Precision + Recall} \quad [6]$$

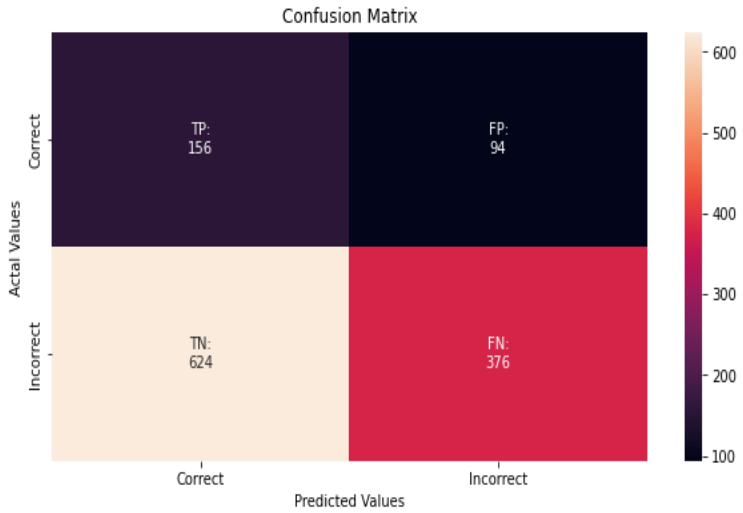
Receiver Operating Characteristic Curve (ROC) is used to illustrate the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across different threshold settings. İt is a graphical representation of the plot between the sensitivity TPR and specificity FPR

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

3. Results and Discussion

In this section, the comparison between the recognition accuracy of Artificial Neural Network (ANN) and K-Nearest Neighbors (KNN) algorithms on voice recognition is presented. The study utilized noisy and noiseless voice datasets, and the performance of both algorithms was evaluated and summarized using figures and tables. It is important to note that the False Acceptance Rate (FAR) and False Rejection Rate (FRR) can be influenced by the threshold set for decision-making. Threshold in the contest of biometric identification is a critical parameter that determines the decision boundary for accepting or rejecting a recognition attempt (Balogun *et al.,* 2022). When a biometric system makes a recognition attempt, it generates a similar score or distance measure between the input biometric data (e.g., voice sample) and the stored template(s) in its database. Score represents the degree of similarity or dissimilarity between the input data and the reference templates (Dini & Saponara, 2021). The threshold value is then used to compare this score against a predefined threshold. Adjusting the threshold involves finding an optimal balance between security and convenience, as highlighted by Shopon *et al.,* (2021).

Furthermore, the study visually presents the prediction performance of ANN and KNN through confusion matrices displayed in Figures 3a and b, 4a and b for ANN on noise and noiseless datasets respectively, and Figures 5a and b, 6a and b for KNN on noise and noiseless datasets respectively. These confusion matrices serve as structured grids to assess the efficacy of the classification models by comparing their predictions against the known true values within the test datasets. They provide valuable insights into the accuracy and reliability of the models, particularly in identifying misclassifications. Each cell in the matrix represents the number of instances that were classified into a particular category (e.g., correctly recognized as a genuine speaker or incorrectly classified as an impostor) compared to the true labels in the dataset. İn the contest of this work, the matrices were used to determine the performance of the considered models (ANN and KNN) on speakers' recognition under varied noisy environments. Overall, this study offers a comprehensive understanding of the performance differences between ANN and KNN algorithms in voice recognition tasks by shedding light on their strengths and limitations in various environmental conditions.



Figures 3a: Noisy ANN Confusion Matrix

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

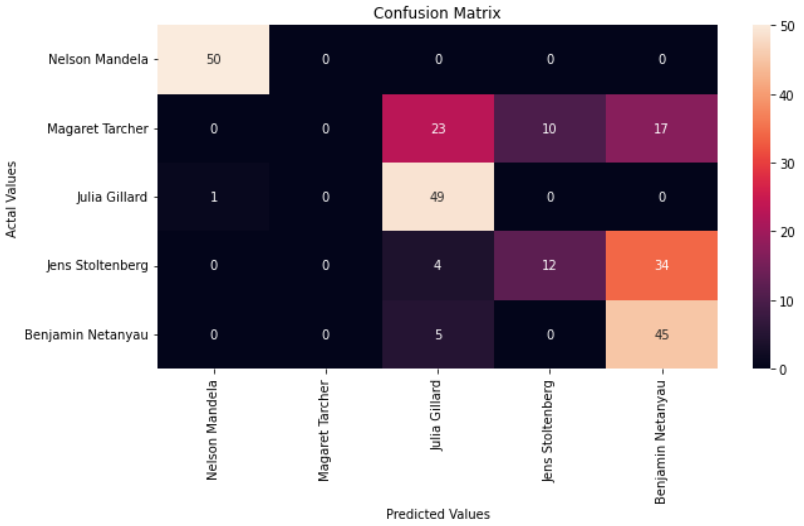


Figure3b: Noisy ANN Speaker Confusion Matrix

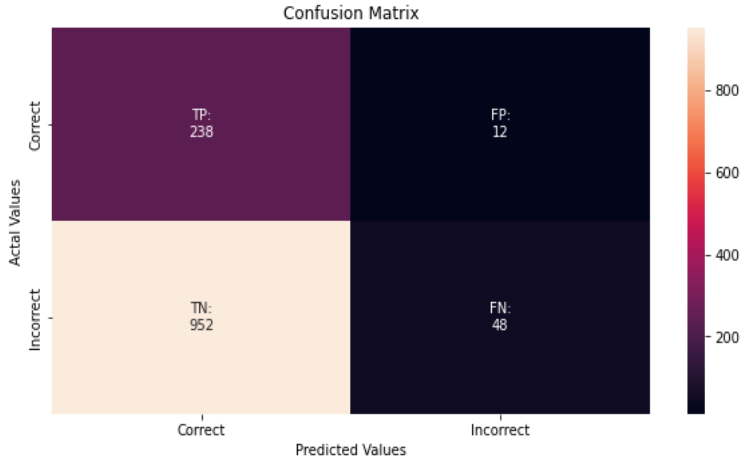


Figure 4a:. Noiseless ANN Confusion Matrix

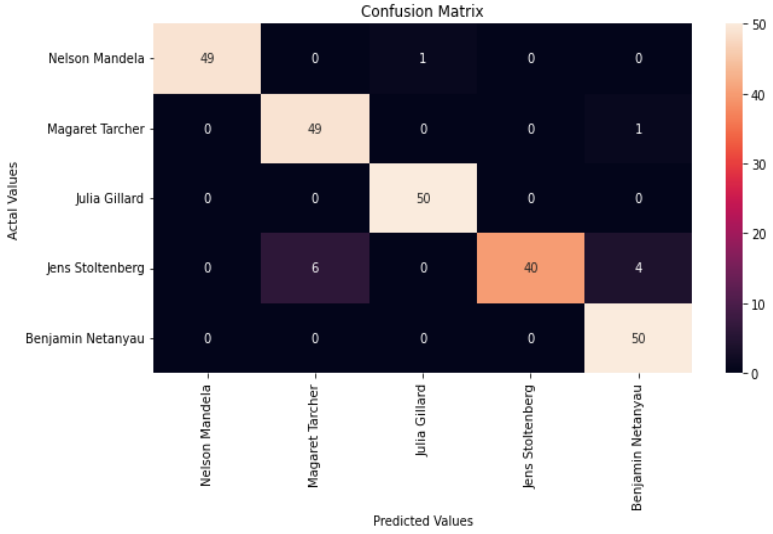


Figure 4b: Noiseless ANN Speaker Confusion Matrix

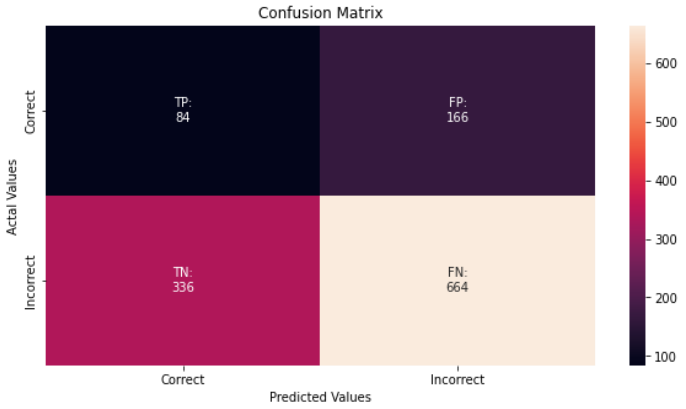


Figure 5a: Noisy KNN Confusion Matrix

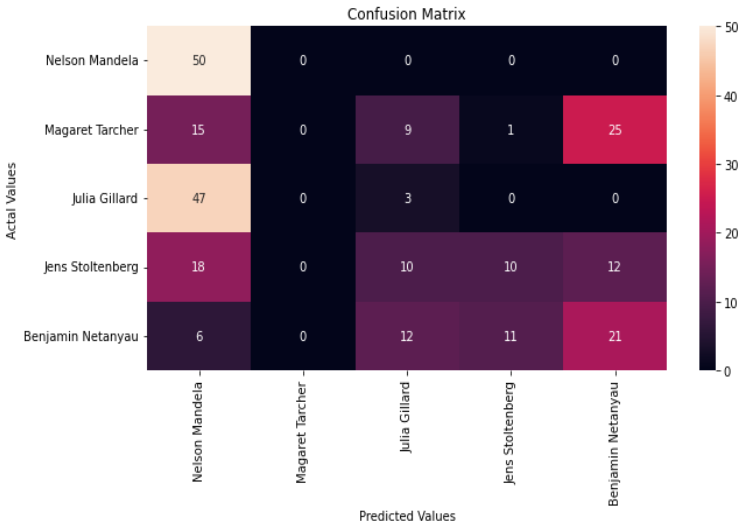


Figure 5b: Noisy KNN Speakers Confusion Matrix

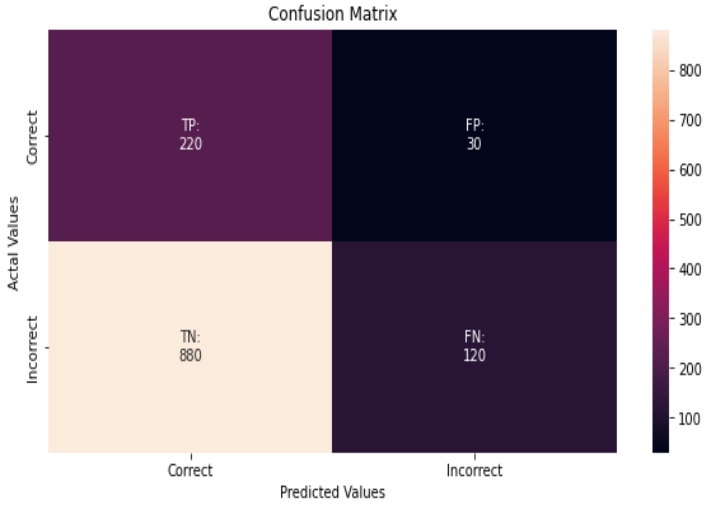


Figure 6a: Noiseless KNN Confusion Matrix

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

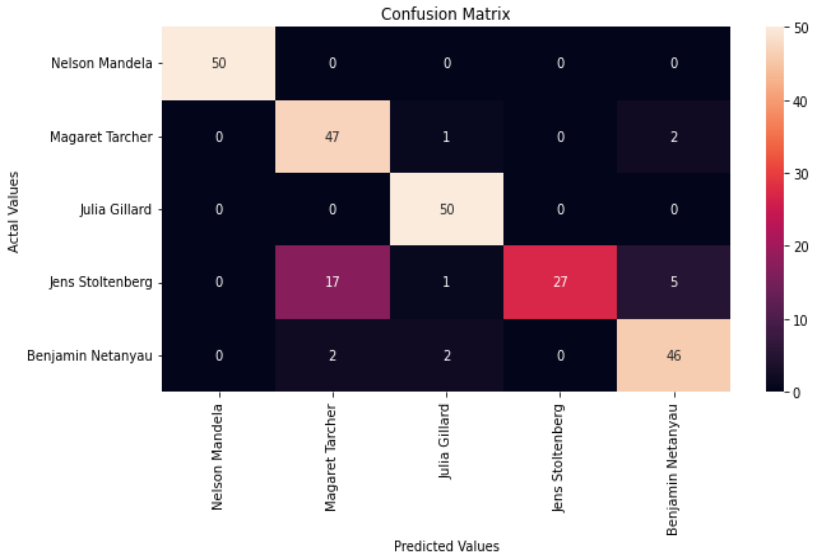


Figure 6b: Noiseless KNN Speaker Confusion Matrix

Table 1: Performance of ANN and KNN on Noisy Validation data

Metrics	Scores% (ANN)	Score% (KNN)
FAR	13.09	33.60
FRR	70.60	88.70
Precision	62.40	33.60
Recall	29.00	11.00
F1_Score	39.00	16.80

Table 1 illustrates the performance of ANN and KNN on noisy data. It reveals that ANN achieved a significantly higher recognition accuracy of 62.40% compared to KNN's accuracy of 33.60% on noisy data. This indicates that ANN outperformed KNN in accurately recognizing speakers, particularly in challenging, noisy environments. It can be seen from the table that ANN generated FAR, FRR, Precision, Recall and F1 scores of 13.09%, 70.60%, 62.40%, 29.00% and 39.00%, respectively, while the corresponding values generated by KNN are 33.60, 88.70, 33.60, 11.00% and 16.80% respectively. According to (Balogun et al., 2023) the lower the FAR and FRR generated by a model the more accurate prediction made by the model. A higher recall indicates that the biometric system is better at capturing and correctly identifying a larger proportion of genuine users. Recall is crucial for ensuring that legitimate users are not wrongly rejected or excluded from accessing the system.

Table 2: Performance of ANN and KNN on Noiseless validation data

Metrics	Scores % (ANN)	Score% (KNN)
FAR	12.00	32.00
FRR	16.70	35.00
Accuracy	95.20	88.00
Precision	95.20	88.00
Recall	83.20	64.70
F1_Score	88.80	74.00

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

Table 2 shows the general performance of ANN and KNN on noiseless dataset. It can be seen from the table that, the recognition accuracy of both algorithms really increased. However, ANN still generated the higher recognition accuracy value of 95.20% compared to that of KNN that is 88.00%. The values of FAR, FRR, Precision, Recall and F1 score generated by ANN are 12.00%, 16.70%, 95.20%, 83.20% and 88.80%, respectively, while that of KNN are 32.00%, 35.00%, 88.00%, 64.70% and 74.00%, respectively. Comparison between Tables 1 and 2 shows that ANN has better recognition accuracy than KNN in all data instances (Noisy and Noiseless).

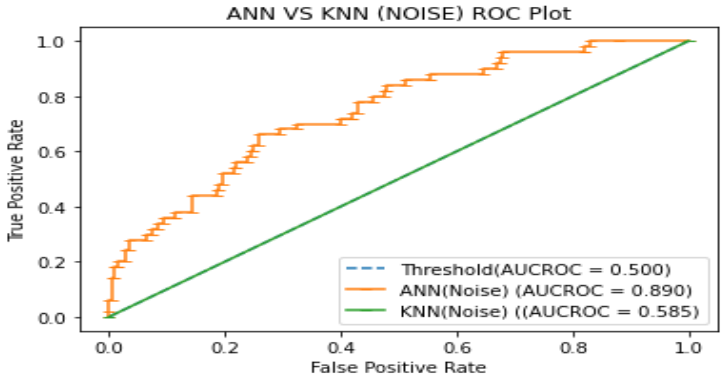


Figure 7: ROC for ANN and KNN on Noisy Validation Data

Present in figure 7 is the Receiver Operating Characteristic (ROC) curve for ANN and KNN. ROC is a graphical representation that illustrates the performance of a binary classification model across different classification thresholds. The ROC curve is a valuable tool in assessing the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate (1-Specificity) at various decision thresholds (Dhole et al., 2023). It can be seen from Figure 7 that the ANN has larger area under the curve than the KNN, The KNN is shown to be on the same line with the threshold. KNN value only surpassed the threshold with a value of 0.085%, hence, it is graphically shown to be laying on the threshold region. This reveals that KNN has a generally poor performance in recognizing noisy data. ANN ability to generate a high true positive rate (accurate recognition) with respect to change in the thresholds even at noisy instances is shown by the sharp edges that make the curve. This implies that ANN is not really affected by noise, or errors usually introduced during capturing stage of biometric system, which is one of the best qualities of good recognition model.

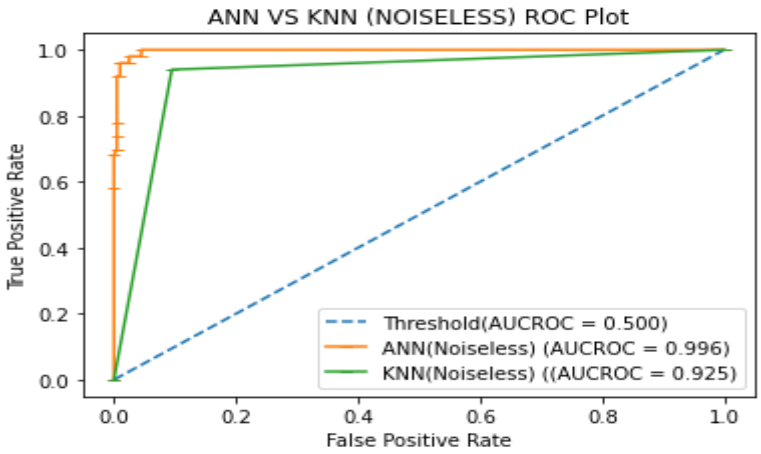


Figure 8: ROC for the ANN and KNN on noiseless validation data

Shown in Figure 8 is the ROC for the performance of ANN and KNN in noiseless dataset. It can be seen from the figure that the recognition accuracy of the two algorithms generally increased with the ANN being near perfect as the area under its curve is about 99.6%, while that of KNN is 92.5%. The ANN curve being closer to True Positive Rate reveals how positive

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

performance it gives with respect to change in the threshold value. The high true positive recognition generated by ANN is an indication that it is more versatile than the KNN.

4. Conclusion

This study evaluated the performance of Artificial Neural Networks (ANN) and K-Nearest Neighbors (KNN) on voice recognition using a Kaggle dataset that includes speech voice data from prominent speakers. The dataset presented challenges with background noise, enhancing the difficulty of the speaker recognition tasks (Li *et al.,* 2024). The pre-processing involved extracting Mel Frequency Cepstral Coefficients (MFCC) from the audio data using the Librosa Python library. KNN utilized time and frequency domain features, while ANN focused solely on MFCCs, known for capturing the formants of audio, particularly human voice. The developed biometric voice recognition system consists of steps such as Speech Sample, Audio Signal Processing, Feature Extraction, Speaker Model Database, Classification Stage, and Evaluation. The ANN implementation involved data preparation, initializing neural network values, backpropagation, calculating outputs of hidden layer neurons, and updating weights iteratively until accuracy improved and potential overfitting avoided, this is in line with (Abushariah *et al.,* 2016). For KNN, the scikit-learn library was used, with parameters like n-neighbors and distance metric set to determine the optimal k-value and metric through heuristics like cross-validation.

The evaluation involved confusion matrices for both ANN and KNN on noisy and noiseless datasets. Various performance metrics, including True Positive, True Negative, False Positive, False Negative, False Acceptance Rate, False Rejection Rate, Accuracy, Precision, Recall, F1 Score, and Receiver Operating Characteristic Curve, were used to compare the algorithms. The results, presented through confusion matrices and performance metrics, provide insights into the effectiveness of ANN and KNN in voice recognition. Specifically, on the noisy dataset, ANN demonstrated recognition accuracy, precision, and recall values of 62.40%, 62.40%, and 29.00%, respectively, while KNN achieved 33.60%, 33.60%, and 11.00%, respectively. On the noiseless dataset, ANN exhibited recognition accuracy, precision, and recall values of 95.20%, 95.20%, and 83.20%, respectively, whereas KNN attained 88.00%, 88.00%, and 64.70%, respectively. Overall, the study discovered that ANN outperformed KNN in terms of recognition, accuracy, precision, and recall on both noisy and noiseless datasets. This observation aligns with previous research by Murthy *et al.*, (2015), which highlights the superior performance of ANN in learning complex patterns, scalability, feature learning capabilities, and robustness to noise, especially in large-scale datasets with high-dimensional feature spaces.

Conclusively, ANN is a better classification algorithm especially when dealing with noisy datasets. The study contributes valuable information to the field of biometric voice recognition, emphasizing the significance of algorithm choice and dataset characteristics in achieving accurate and reliable results.

Future improvement in the work

The following are recommended for future improvement in the developed systems:

- Hybridization of the two algorithms (ANN and KNN) will improve the overall system performance in biometric systems.
- The data used to evaluate the two considered algorithms in this work were collected in an unrestrained environment. Implementation of the algorithms can also be done using data collected in a controlled environment to see the effect this will have on the algorithm's recognition accuracy.
- Implement data augmentation techniques to artificially increase the size of the dataset. This can help improve the robustness of the models, especially when dealing with limited data.
- Investigate the potential benefits of ensemble methods, such as combining multiple classifiers to enhance overall performance. Ensemble techniques like stacking or boosting could be explored.

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

REFERENCE

Abushariah, M., Journal, A. A.-E. S., & 2016, . Automatic Identity Recognition Using Speech Biometric. Core.Ac.Uk, 12(12): 1857–7881. https://doi.org/10.19044/esj.2016.v12n12p43

Alattar, Z. Sh., Abbes, T., & Zerai, F. 2023. Privacy-preserving hands-free voice authentication leveraging edge technology. Security and Privacy, 6(3): e290. https://doi.org/10.1002/spy2.290

Alharbi, B., & Alshanbari, H. S. 2023. Face-voice based multimodal biometric authentication system via FaceNet and GMM. PeerJ Computer Science, 9:1468. https://doi.org/10.7717/peerj-cs.1468

Al-Zakarya, M., & Al-Irhaim, Y. 2023. Unsupervised and Semi-Supervised Speech Recognition System: A Review. AL-Rafidain Journal of Computer Sciences and Mathematics, 17(1): 34-42. https://doi.org/10.33899/csmj.2023.179466

Ammour, N., Jomaa, R. M., Islam, M. S., Bazi, Y., Alhichri, H., & Alajlan, N. 2023. Deep Contrastive Learning-Based Model for ECG Biometrics. Applied Sciences (Switzerland), 13(5):3070. https://doi.org/10.3390/app13053070

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G. and Chen, J., 2016. Deep speech 2: End-to-end speech recognition in english and mandarin. International conference on machine learning:173-182: PMLR.

Anil, B.V., and M. S. Ravikumar 2022. Biometric Recognition from Face-Voice Using Rough-Neuro-Fuzzy Classifiers. In High Performance Computing and Networking: Select Proceedings of CHSN 2021, Singapore: Springer Singapore, 2022, 489-501. https://doi.org/10.1007/978-981-16-9885-9_40

Balogun, M. O., Adeola Odeniyi, L., Olusola Omidiora, E., Olabiyisi, S. O., & Falohun, A. S. 2022. Optimized Negative Selection Algorithm for Image Classification in Multimodal Biometric System. Acta Informatica Pragensia, 12(1): 3-18 https://doi.org/10.18267/j.aip.186

Benmachiche, A., Hadjar, B., Boutabia, I., Betouil, A.A., Maatallah, M. and Makhlouf, A., 2022. Development of a biometric authentication platform using voice recognition. 2022 4th IEEE. International Conference on Pattern Analysis and Intelligent Systems (PAIS): 1-7. https://doi.org/10.1109/PAIS56586.2022.9946890

Booysens, A. and Viriri, S., 2022. Exploration of Ear Biometrics using EfficientNet. Computational Intelligence and Neuroscience, 2022(1): 3514807. https://doi.org/10.1155/2022/3514807

CAN, Z., & ATILGAN, E. 2023. A Review of Recent Machine Learning Approaches for Voice Authentication Systems. Bilgi ve İletişim Teknolojileri Dergisi, 5(1): 96-114. https://doi.org/10.53694/bited.1296035

Collobert, R., Hannun, A. and Synnaeve, G., 2019. A fully differentiable beam search decoder. International Conference on Machine Learning, PMLR: 1341-1350. https://proceedings.mlr.press/v97/collobert19a.html

Collobert, R., Puhrsch, C. and Synnaeve, G., 2016. Wav2letter: an end-to-end convnet-based speech recognition system. arXiv preprint arXiv:1609.03193. https://arxiv.org/abs/1609.03193

Demidova, L.A., 2021. Two-stage hybrid data classifiers based on SVM and kNN algorithms. Symmetry, 13(4): 615. https://doi.org/10.3390/sym13040615

Dhole, S.A., Patil, J.K., Jagdale, S.M., Reddy, H.G. and Gowda, V.D., 2023. Multimodal Biometric Identification system using Random Selection of Biometrics. International Journal of Electrical and Electronics Engineering, 10: 63-73.

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

https://doi.org/10.14445/23488379/IJEEE-V10I1P106
Dini, P. and Saponara, S., 2021. Analysis, design, and comparison of machine-learning techniques for networking intrusion detection. Designs, 5(1): 9.
https://doi.org/10.3390/designs5010009
Dvořák, M., Drahanský, M. and Abdulla, W.H., 2021. On the fly biometric identification system using hand-geometry. IET Biometrics, 10(3): 315-325.
https://doi.org/10.1049/bme2.12024
Graves, A., Fernández, S., Gomez, F. and Schmidhuber, J., 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd international conference on Machine learning: 369-376.
https://doi.org/10.1145/1143844.1143891
Gupta, V., Garade, S., Kothekar, S., Meshram, D., Wankhede, S., & Pote, Prof. R. 2023. An Investigative Study of Voice Functioned Smart Door Lock System. International Journal of Research Publication and Reviews, 04(01): 122-131.
https://doi.org/10.55248/gengpi.2023.4154
Hannun, A., Lee, A., Xu, Q. and Collobert, R., 2019. Sequence-to-Sequence Speech Recognition with Time-depth Separable Convolutions. arXiv preprint arXiv:1904.02619.
https://doi.org/10.21437/Interspeech.2019-2460
He, J., He, J. and Zhu, L., 2020, June. Text-independent speaker verification based on triplet loss. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (1): 2385-2388.
https://ieeexplore.ieee.org/abstract/document/9085125/
Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6): 82-97.
https://doi.org/10.1109/MSP.2012.2205597
Hitelman, A., Edan, Y., Godo, A., Berenstein, R., Lepar, J. and Halachmi, I., 2022. Biometric identification of sheep via a machine-vision system. Computers and Electronics in Agriculture, 194: 106713.
https://doi.org/10.1016/j.compag.2022.106713
Imdad, U., Ahmad, W., Asif, M. and Ishtiaq, A., 2017, December. Classification of students results using KNN and ANN. IEEE 2017 13th international conference on emerging technologies (ICET), 1-6.
https://doi.org/10.1109/ICET.2017.8281651
Jawalkar, P.A., Akshaya, B.S., Sraveena, B. and Sahithi, B.S., 2023. Object Detection and Multi-class Classification from Videos: An Application of AI-enabled Surveillance Camera. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 14(03): 904-910.
https://doi.org/10.22214/ijraset.2023.53375
Juang, B.H. and Rabiner, L.R., 1991. Hidden Markov models for speech recognition. Technometrics, 33(3): 251-272.
https://www.tandfonline.com/doi/abs/10.1080/00401706.1991.10484833
Kaensar, C., 2013. A comparative study on handwriting digit recognition classifier using neural network, support vector machine and k-nearest neighbor. In The 9th International Conference on Computing and InformationTechnology (IC2IT2013) 9th-10th May 2013 King Mongkut's University of Technology North Bangkok, 155-163.
https://doi.org/10.1007/978-3-642-37371-8_19
kaggle dataset for speech recognition Kaggle
https://www.kaggle.com › datasets › tags=16683-Auto...

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

Kao, C.Y. and Chueh, H.E., 2023. Voice Response Questionnaire System for Speaker Recognition Using Biometric Authentication Interface. Intelligent Automation & Soft Computing, 35(1). https://doi.org/10.32604/iasc.2023.024734

Ko, T., Peddinti, V., Povey, D., Seltzer, M.L. and Khudanpur, S., 2017, March. A study on data augmentation of reverberant speech for robust speech recognition. IEEE 2017 International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5220-5224. https://ieeexplore.ieee.org/abstract/document/7953152/

Lanjewar, R.B., Mathurkar, S. and Patel, N., 2015. Implementation and Comparison of Speech Emotion Recognition System using Gaussian Mixture Model (GMM) and K-Nearest Neighbor (K-NN) Techniques. Procedia computer science, 49: 50-57. https://www.sciencedirect.com/science/article/pii/S1877050915007358

Li, K., Baird, C. and Lin, D., 2023. Defend Data Poisoning Attacks on Voice Authentication. IEEE Transactions on Dependable and Secure Computing. https://doi.org/10.1109/TDSC.2023.3289446

Li, Q., Mai, Q., Wang, M. and Ma, M., 2024. Chinese dialect speech recognition: a comprehensive survey. Artificial Intelligence Review, 57(2): 25. https://doi.org/10.1007/s10462-023-10668-0

MK, A.H.K. and Aithal, P.S., 2022. Voice Biometric Systems for User Identification and Authentication–A Literature Review. International Journal of Applied Engineering and Management Letters (IJAEML), 6(1): 198-209. https://doi.org/10.47992/ijaeml.2581.7000.0131

Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M. and Zhang, D., 2023. Biometrics recognition using deep learning: A survey. Artificial Intelligence Review, 56(8): 8647-8695.). https://doi.org/10.1007/s10462-022-10237-x

Moosavian, A., Ahmadi, H., Tabatabaeefar, A. and Khazaee, M., 2013. Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing. Shock and Vibration, 20(2): 263-272. https://doi.org/10.3233/SAV-2012-00742

Murthy, H.N. and Meenakshi, M., 2015. ANN, SVM and KNN classifiers for prognosis of cardiac ischemia-a comparison. Bonfring International Journal of Research in Communication Engineering, 5(2): 7. https://doi.org/10.9756/BIJRCE.8030

Nigam, D., Patel, S.N., Raj Vincent, P.D., Srinivasan, K. and Arunmozhi, S., 2022. [Retracted] Biometric Authentication for Intelligent and Privacy-Preserving Healthcare Systems. Journal of Healthcare Engineering, 2022(1): 1789996. https://www.hindawi.com/journals/jhe/2022/1789996/

Pudurkar, R., Patil, S., Ansari, G., & Phiroj, S. (2022). Biometric Voice Recognition system using MFCC and GMM with EM. *International Journal of Computer Applications*, *184*(26): 21-30. https://doi.org/10.5120/ijca2022922315

Richardson, F., Reynolds, D. and Dehak, N., 2015. Deep neural network approaches to speaker and language recognition. IEEE signal processing letters, 22(10): 1671-1675. https://doi.org/10.1109/LSP.2015.2420092

Saon, G., Soltau, H., Nahamoo, D. and Picheny, M., 2013, December. Speaker adaptation of neural network acoustic models using i-vectors. 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 55-59.

Shaw, P., Uszkoreit, J. and Vaswani, A., 2018. Self-attention with relative position representations. arXiv preprint arXiv: 1803.02155.

.

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng

https://doi.org/10.18653/v1/n18-2074
Snyder, D., 2020. X-Vectors: Robust neural embeddings for speaker recognition (Doctoral dissertation, Johns Hopkins University).
Tanberk, S. and Tükel, D.B., 2022, July. Ensemble Learning with CNN–LSTM Combination for Speech Emotion Recognition. In Proceedings of International Conference on Computing and Communication Networks, Singapore: Springer Nature Singapore: ICCCN 2021, 39-47. https://doi.org/10.1007/978-981-19-0604-6_5
Wang, T.Y. and Kumar, A., 2016, February. Recognizing human faces under disguise and makeup. In 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), 1-7. https://ieeexplore.ieee.org/abstract/document/7477243/
Wendehorst, C. and Duller, Y., 2021. Biometric Recognition and Behavioural Detection Assessing the ethical aspects of biometric recognition and behavioural detection techniques with a focus on their current and future use in public spaces. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4087455
Xu, Z., Zhang, J., Zhang, Y., Zhou, M., Li, K., Geng, S. and Zhang, X., 2020. Stroke controllable style transfer based on dilated convolutions. IET Computer Vision, 14(7): 505-516. https://doi.org/10.1049/iet-cvi.2019.0912
Xue, J. and Zhou, H., 2023. Physiological-physical feature fusion for automatic voice spoofing detection. Frontiers of Computer Science, 17(2): 172318. https://doi.org/10.1007/s11704-022-2121-6

Corresponding author's e-mail address: monsurat.balogun@kwasu.edu.ng