http://dx.doi.org/10.14500/aro.11897

Improved Kurdish Dialect Classification Using Data Augmentation and ANOVA-Based Feature Selection

Karzan J. Ghafoor, Sarkhel H. Taher, Karwan M. Hama Rawf and Ayub O. Abdulrahman

Computer Science Department, College of Science, University of Halabja,
Halabja, 46018, Kurdistan Region - F.R. Iraq

***Abstract*—Analyzing dialects in the Kurdish language proves to be tough because of the tiny phonetic distinctions among the dialects. We applied advanced methods to enhance the precision of Kurdish dialect classification in this research. We examined the dataset's stability and variation through the use of time-stretching and noise-augmenting methods. Analysis of variance (ANOVA) filter approach is applied to improve feature selection (FS) more efficiently and highlight the most relevant features for dialect classification. The ANOVA filter method ranks features based on the means from different dialect groups, which made FS better. To make dialect classification work better, a 1D convolutional neural network model was given a dataset that had ANOVA FS added to it. The model showed a very strong performance, reaching a remarkable accuracy of 99.42%. This noteworthy increase in accuracy beat former research with an accuracy of 95.5%. The findings demonstrate how combining time stretch and FS methods can improve the accuracy of Kurdish dialect classification. This project improves our understanding and implementation of machine learning in the field of linguistic diversity and dialectology.**

***Index Terms*—1D convolutional neural network, Data augmentation, Feature selection, Kurdish dialect identification, Sound features.**

I. Introduction

The rapid expansion of voice recognition technology is supported by progress in machine learning and the extensive use of the Internet. Communities with popular languages, including Chinese and English, have developed rigorous access to considerable data. Before we can achieve automatic recognition, we must address the Kurds' limited resources and lack of an extensive open corpus. Due to the small number of experts studying Kurdish language recognition, there is a lack of a robust research foundation that limits many advanced voice recognition techniques. Kurdish study is straightforward due to its dialects' easy clustering and linear separation (Ghafoor, et al., 2021). Numerous languages use the Arabic script as a main popular way to write in the Middle East and North Africa. Numerous languages use it greatly, including Arabic and Urdu. The Kurdish language includes 34 characters. Since Arabic is the root script for the Kurdish alphabet, the Arabic script includes the 23 most frequently used letters from Arabic (Hama Rawf, Abdulrahman and Mohammed, 2024). Furthermore, the Arabic script shares similarities with the languages and writing systems of many Middle Eastern nations, which could facilitate education. Four central Middle Eastern states have their own dialects, where Kurds speak their language. The Kurdistan region of Iraq contains three main dialects of Kurdish language including Sorani and Badini as well as Hawrami which are the most widely used vernaculars. These dialects serve as the main spoken language forms among Kurdish groups who live in Iraq where they maintain their own linguistic features. Languages can achieve better accuracy in speech processing through models specifically made for different dialectal variations because dialect classification provides essential knowledge for language recognition research. Communities speaking Kurdish reside in territories of Turkey, Iraq, Iran, and Syria. The current estimates indicate that more than 40 million converse in Kurdish. Both the recognition of KuSL and dialect recognition systems (DRS) facilitate better communication for individuals with language diversity. The many influences on speech recognition are investigated by DRS, and strategies are offered for merging them with systems for recognizing various dialects (Rawf, et al., 2024). The Kurdish language lacks large enough datasets, due to its complexity and variety of dialects, making it tough to develop efficient natural language processing (NLP) tools, especially for Kurdish Named Entity Recognition (KNER). To solve this, Abdullah et al. (2024) adapted a pre-trained RoBERTa model specifically for KNER. This included creating a new Kurdish corpus, adjusting the model setup, and fine-tuning the training process. Results showed that using sentence-piece tokenization, they improved the F1 score by 12.8%, setting new standards for Kurdish NLP applications. Automatic speech recognition applications have incorporated data augmentation (DA) techniques to generate more training data, thereby improving data quality in terms of both quantity and diversity. As a result of this approach, the system gains strength and avoids overfitting (Nguyen, et al., 2020). Kanda, Takeda and

ARO-The Scientific Journal of Koya University
Vol. XIII, No. 1 (2025), Article ID: ARO.11897. 10 pages
DOI: 10.14500/aro.11897
Received: 14 November 2024; Accepted: 27 February 2025
Regular research paper; Published: 07 March 2025
†Corresponding author's e-mail: karzan.ghafor@uoh.edu.iq

Obuchi (2013), and Ragni et al. (2014) talk about how to use unsupervised learning and fake data to improve model training in places with few resources. DA, in conjunction with feature selection (FS) algorithms, reduces overfitting. This method enhances the performance of speech recognition systems significantly and outperforms various other alternatives (Sangwan, Deshwal and Dahiya, 2021). The FS and extraction of the source voice signal must yield maximum recognition effectiveness while keeping computational needs low. Identify data a machine learning classifier uses input features pulled from the initial signals (Abdul, Al-Talabani and Abdulrahman, 2016). To handle the problem created by the small corpus, a variety of DA methods, including temporal stretching, noise injection, and audio amplification, was employed. The created dataset will educate acoustic and language models (Lounnas, Lichouri and Abbas, 2022). Among all techniques for voice recognition, other than DA, it provides the biggest increase in effectiveness. The analysis from Peddinti et al. (2015) indicates that implementing DA increased efficacy by 33%. The comprehensive review set features both important and irrelevant characteristics for the targeted classifiers, lowering their effectiveness. To uncover the key elements and discard unimportant information, it is crucial to implement FS routines. By employing FS approaches, it is possible to select the optimal and important features that improve the identification method. In addition, the K-fold cross-validation method may be applied to guarantee trustworthy outcomes and lower the chances of overfitting (Tubishat, et al., 2019). To correctly classify dialects of the Kurdish language in this study, researchers use DA and FS methods derived from convolutional neural network (CNNs) and employ the shortest audio segment. This study seeks to assess how effective DA and FS are. The work comes up with a creative way to group dialects using CNN to choose which features to use. Using a hybrid approach that uses a CNN to show the unique features of each dialect makes it more effective at recognizing them. This approach simplifies the requirement for high-cost human feature identification. This innovative approach provides a functional solution for identifying and classifying dialects in varied linguistic and cultural situations. The following structure guides the subsequent sections of the paper: Section 2 summarizes existing literature and research concerning the topic. Section 3 presents the materials and techniques used in the study. In Section 4, details of the DA mechanism are provided. The FS appears in section five. In Section 6 you can find an account of results along with discussion, and Section 7 includes the closure and prospective endeavors.

II. Related Work

Different ways of identifying dialects show differences that depend on a number of factors, such as the language being studied and the features that are extracted. Investigators employed machine-learning algorithms with multiple techniques for feature extraction and selection to efficiently sort dialects in various languages. The conversion of Gaussian mixture model (GMM) weights included an analysis implemented by a positive factor analysis technique (Bahari, et al., 2014). Based on their study, GMM loads serve as extra data for GMM revealing DR and language but collectively provide less data. Among Bangladeshi dialects, utilizing GMM to define characteristics is the Mel Frequency Cepstral Coefficient (MFCC), along with its Delta and Delta-delta. Astonishingly built from the sound of utterances (Das, et al., 2016). Moreover, Mulahuwaish et al. (2020) propose an effective system for the web that gathers news articles into four distinct groups: Business, technology and science, health, and entertainment. Researchers assess four different machine learning classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (kNN), decision tree, and long short-term memory. The classifiers are applied individually and then assessed using accuracy and receiver operating characteristic curves. The results show that while SVM attained the maximum accuracy at 95.04%, kNN attained the lowest accuracy at 88.72%.

By changing the structure of the training dataset, DA makes alternative training sets. Various study areas, such as image processing and sound classification, find DA to be a useful technique (Li, et al., 2020). In many research domains, the DA technique is extensively used. A small number of samples from several DA classes has a big effect on how problems are solved in a broad classification setting (Zheng, et al., 2020). For speaker identification, a strategy named DA-DNN7L is available to increase data samples. The method uses white noise augmentation and time manipulation procedures. Using a deep neural network (DNN) enables the techniques to develop a new model. Increasing the available data from speakers of Indonesian ancestry is aimed at using a data enhancement approach. When assessed against different multilayer approaches, the seven-layer DNN yields fair accuracy. In addition, the use of the most effective seven-layer DNN DA approach in research yielded an accuracy rate of 99.76% and a loss of 0.05 when using a 70%: The ratio was 30%, 400 data points were added, and a CNN and DA method were used to accurately find sounds that were close by (Nugroho and Noersasongko, 2022). The results showed that combining a deep, high-capacity system with an increased training set yields better outcomes. The fusion turns a better result than the advised CNN without augmentation and bottom models relying on dictionary learning with augmentation. The study (Salamon and Bello, 2017) recorded an accuracy of 94%.

A proposal exists for a speaker identification approach tailored to a special wearable tool for reducing gender-based violence. A detailed study was implemented to gauge how stress affects speech with DA methods. Findings demonstrate that with naturally stressed samples in the training set, efficacy is satisfactory. When these missing components are unavailable, we can improve the results by introducing artificially generated stress-like samples (Rituerto-González, et al., 2019).

To choose the most important features and remove unnecessary data, one can implement algorithms. Determining the essential features greatly depends on these algorithms and leads to an improved performance in sentiment analysis. The three types of FS methods include filter methods, wrapper approaches, and hybrid methods. The analysis of features through filter approaches resembles established

strategies for FS (Tubishat, et al., 2019). To achieve high recognition efficiency while reducing computational resources substantially, the input features and extraction method for the voice signal should be optimized (Sangwan, Deshwal and Dahiya, 2021). A novel FS method emerged from fusing the Binary Bat technique with late acceptance hill-climbing. This algorithm targets the selection of important feature vectors that will lessen model complexity and speed up training. Using the Indian TTS dataset made by IIT-Madras, the Random Forest algorithm hits an efficiency of 92.35% (Das, et al., 2020).

Speaker verification is achieved through a super vector that features the mean values of each phonetic sound. The creation of the first accent models is through the calculation of averages for each class using speech super vectors. The methods evaluate a Flemish audio dataset that classifies speakers into five groups. Selecting a subset of features results in an over 20% improvement in accessing rate across all models. Wu et al. (2010) say that using speaker vectors during the creation phase gets much better results than a regular GMM that works directly with the main feature vectors in both text-dependent and text-independent situations. In the audio classification process, FS and combination are essential since they can boost the performance of deep learning models significantly. Three datasets were evaluated by several advanced deep-learning models to assess their functionality. The findings show that the features chosen are affected by both the dataset and the model chosen (Turab, et al., 2022).

An important factor in achieving impressive outcomes was the use of DA and FS. A novel approach to identify emotions in speech (SER) is shared that incorporates data enlargement procedures and feature combination along with their selection. The studies are conducted using two available datasets: The IEMOCAP database along with the Chinese Hierarchical Speech Emotion Dataset of Broadcasting (DB) constitute two datasets. According to the study (Tu, et al., 2023), the given framework has test accuracies of 66.44% and 93.47% for the unweighted typical basis. A technique for identifying emotions in spoken language was created to solve the problems in currently available approaches and apply it to Arabic speech. The model applies data enhancement techniques while supplying the predetermined features to a transformer model for emotion recognition. Four datasets made up the evaluation: BAVED along with emotional database (EMO)-DB and SAVEE as well as EMOVO. For these datasets, the results showed accuracies of 95.2%, 93.4%, 85.1%, and 91.7% (Al-onazi, et al., 2022).

III. Materials and Methodology

For dialect recognition classification purposes, a CNN model with various feature enhancement techniques such as add_white_noise and time_stretch is utilized. The results produced by these techniques become input for the SelectKBest FS algorithm. They then send those attributes into a revised CNN to analyze. The model implements training and assessment processes while using datasets marked with dialect information. The model can identify dialects by analyzing the collected traits and provides important understanding of the variety in spoken language. The proposed DRS is represented in Fig. 1.

A. Dialectal Speech Dataset

The primary database for this study is the "KuLD" dataset, compiled by researchers in the Computer Science Department at the University of Halabja. Data collection lasted over months. At all phases of gathering data, careful observance of established methods was constant. The dataset included speakers across all age groups and genders. For the three dialects, Sorani, Badini, and Hawrami, a tally of 2000 samples was achieved. Each sample in the dataset runs for 1 s and totals 6000 s (Rawf, et al., 2024).

B. CNN Architecture

The method applied organized Kurdish dialects using a 1-D CNN based on the KuLD database filled with sound information. The dataset utilized in this study has been identified as being comprised of Sorani Badini and Hawrami samples. CNN has received considerable attention in computer vision and audio processing. These fields cover tasks that operate without an exact spatial location in spectrogram visuals. CNN is a quick and easy way to accurately group spectrogram features into different groups (Khamparia, et al., 2019). United FS and data bolstering in CNNs raise accuracy and durability. Training data diversity improves with DA strategies to enhance how models detect patterns as well. Using transfer learning, the network highlights significant features to improve its efficiency. They produce adaptable and reliable models together. With audio DA included, the CNN structure shows sound classification efficiency leading the field (Salamon and Bello, 2017). The proposed 1D CNN architecture consists of twelve layers: a set of input signals is presented, followed by five convolution layers and a MaxPooling stage, as well as three dense layers ending in the output layer. Fig. 2 depicts the structure presented in the 1D CNN model.

Consequently, the parameters of the approach have been meticulously tuned to achieve a significant amount of efficiency in categorizing the Kurdish dialect. The precise specifications of the presented CNN architecture are outlined in Table I.

IV. DA

DA, alternately abbreviated as DA, is a critical technique that researchers employed to increase the size of the dataset by transforming existing data. This strategic maneuver, which has proved beneficial in training neural networks according to Rebai et al. (2017), has considerable implications for deep learning, especially for small data sets, as shown by Ma, Tao and Tang (2019). As an approach, DA can offset overfitting impedance, increase model reliability (Moreno-Barea, Jerez and Franco, 2020), and improve its generality, which is a common issue with machine learning (Ma, Tao and Tang, 2019). When analyzing the vast field of machine learning, the use of DA in the context of the deep learning model assumes

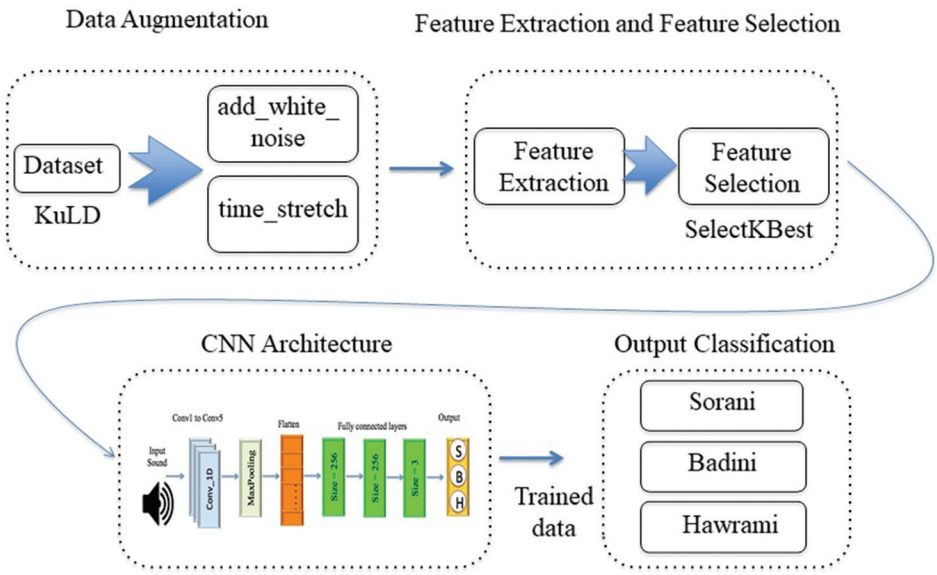

Fig. 1. A block diagram of the proposed model.

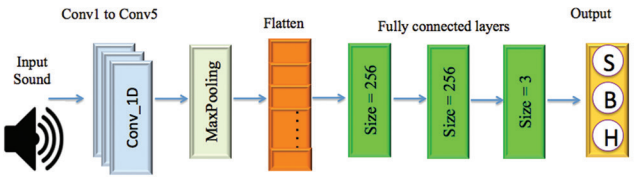

Fig. 2. The comprehensive architecture of a one-dimensional (CNN) architecture.

TABLE I
PRESENTS AN OVERVIEW OF THE DEMONSTRATED MODEL'S STRUCTURE

Layers' type	Output shape	Activation function	Parameter
conv1d_1 (Conv1D)	(None, 182, 128)	relu	640
conv1d_2 (Conv1D)	(None, 179, 64)	relu	32832
conv1d_3 (Conv1D)	(None, 176, 32)	relu	8224
conv1d_4 (Conv1D)	(None, 173, 128)	tanh	16512
conv1d_5 (Conv1D)	(None, 170, 128)	tanh	65664
max_pooling1d_1	(None, 28, 128)	-	0
flatten_1 (Flatten)	(None, 3584)	-	0
dense_1 (Dense)	(None, 256)	sigmoid	917760
dense_2 (Dense)	(None, 256)	sigmoid	65792
dense_3 (Dense)	(None, 3)	softmax	771

a pivotal position when trying to enhance the predictive capabilities of an organization, particularly when utilizing large databases. This principle is true as proved by the study conducted by Moreno-Barea and association (Moreno-Barea, Jerez and Franco, 2020). DA comes in various forms, among which the most common are white noise injection and time stretching (TS), among others that form part of the many strategies used in this branch.

A. White_Noise

One of the main problems of deep learning is the problem associated with the use of small volumes of data. They pointed out that a possible way to solve this problem is by injecting noise during training. Rebai et al. (2017) have proven that the incorporation of white noise provides marketing feedback in speaker recognition. This approach involves the introduction of another noise of random signals of equal amplitudes at different frequencies (Moreno-Barea, Jerez and Franco, 2020). However, it should be pointed out that in the case of audio signals, the frequency range in question can be located substantially in the audible range, which generally varies between 20 Hz and 20 kHz, as shown in Fig. 3.

Therefore, we note that the integration of White Noise shows significant enhancements in the efficiency of the voice recognition models, as noted by Hu, Tan and Qian (2018) and Aguiar, Costa and Silla (2018). This procedure involves joining clean audio with noise, introducing a new sound, which makes it a perfect technique for the DA process, which is a valuable technique in deep learning.

B. Time_Stretching

An audio modification strategy that changes the rate or length of a signal without changing its pitch is referred to as TS. This method is especially helpful in signal filtering for musical signals containing tonal, noise, and transient mix components. Such signals include singing, techno music, and jazz recordings with voices (Damskägg and Välimäki, 2017). TS is a technique used by different studies to conduct DA. Several techniques, such as the synchronous overlap and add algorithm, fuzzy techniques, and CNN, have been used together with TS. These approaches have over and over again boosted the sophistication of the proposed models, as mentioned by Salamon and Bello (2017) in their article and Kupryjanow and Czyzewski (2012).

Voice signal enhancement using the "Time Stretching" directly assists in providing an esthetically represented concept of one of the methods in DA. This is done with the help of making changes in the time stretch, indicated in Fig. 4, while the original voice is analyzed and the sample's time duration is changed.

http://dx.doi.org/10.14500/aro.11897

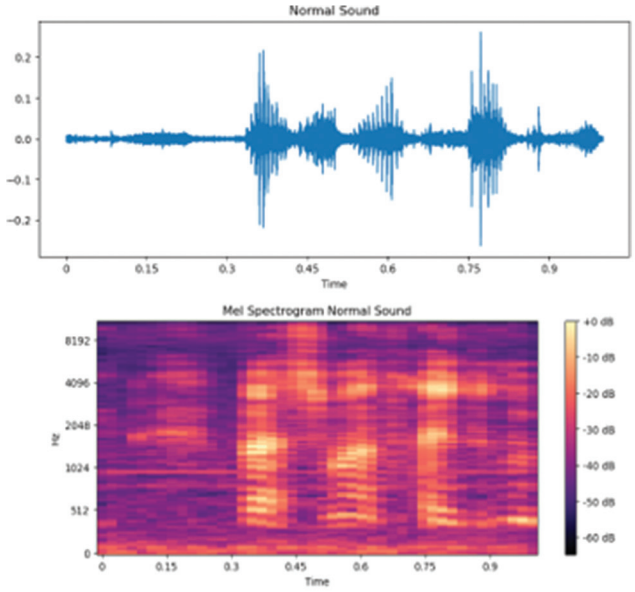

Fig. 3. Normal Sound waveform and Mel_spectogram.

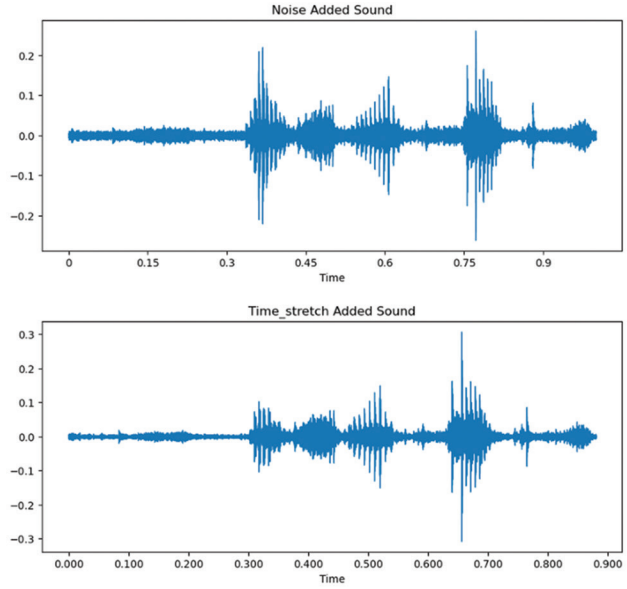

Fig. 4. Noise added and Time_Stretch waveform.

C. Implementation of Analysis of Variance (ANOVA)-Based FS

The ANOVA-based FS method applied through the SelectKBest function in the Scikit-learn library optimized the Kurdish dialect classification feature set. The F-statistic helps the evaluation process identify statistically important features for the classification task.

The first step in extracting an audio dataset was to get four important features: MFCC, Mel Spectrogram, Delta MFCC, and Spectral Contrast. ANOVA evaluated and ranked the combined feature vector, which contained these features. Selection of the most significant features (200) occurred through the ANOVA top-k FS method. The proposed feature extraction method, implemented as Algorithm 1, uses Fig. 5 to represent its procedure through a flowchart.

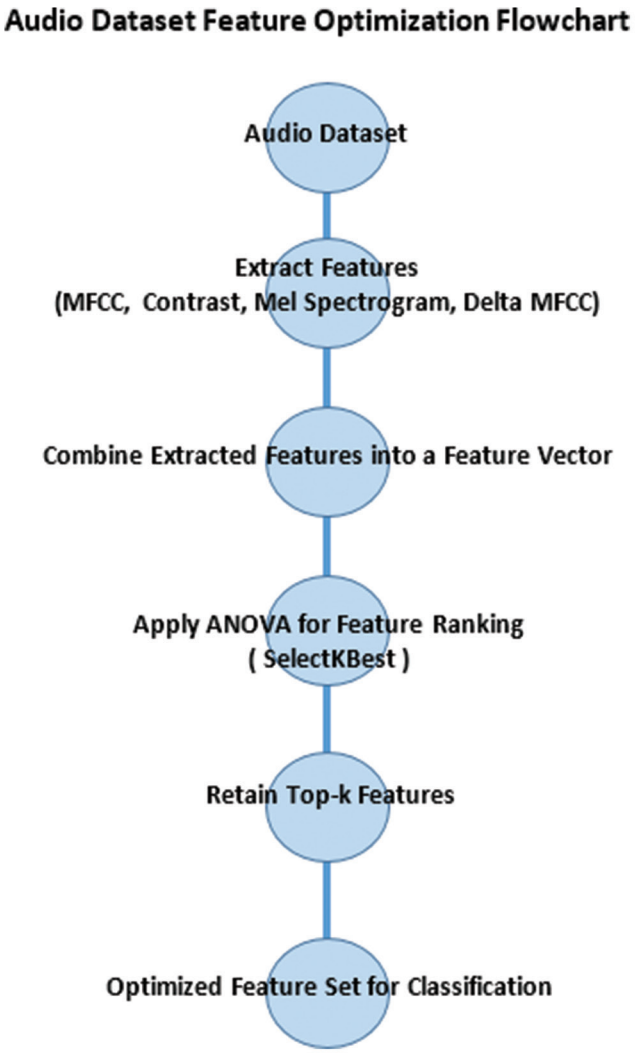

Fig. 5. Flowchart of implementation of analysis of variance-based feature selection.

By applying this approach, it becomes possible to utilize only crucial features that make the model both perform better and require less computational power.

V. FS

It is part of the preprocessing steps that only a few of the most important features in a dataset will be used. This gets rid of features that are not important to the model or are not important at all. In other words, it simplifies the process by removing features from the initial set that may not be essential or useful for defining important features. Numerous FS techniques have been designed specifically to enable choosing valuable data while, at the same time, attempting to exclude data that are insignificant (Shetty, Patnaik and Prasad, 2022) and (Zhou, Wang and Zhu, 2022). It is suggested that it is reasonable to ensure FS's impact on the performance of the outcomes. Sometimes, irrelevant features influence the result; hence, there is a need to either ignore them or consider only the features with the highest rank for the classifier.

This study uses a FS filtering method. After making p-value matrices, an ANOVA test is used to see if there are significant differences between the mean values of a hypothesis-based number of groups. Interestingly, in relation to the F-statistic in this particular setup, they do function to provide some insight regarding feature ranking. When we obtain a higher F-statistic value, it indicates that the identified features are of a higher importance level. In other words, the F-statistic acts as a measure to determine whether the mean of different samples is significantly different. For example, in ANOVA, a feature with an F-statistic value high enough to reject the null hypothesis is needed; the level of significance depends on the F-statistic value (Cheng, et al., 2020). In addition, this approach saves time required to train the classifier because large or small values are not necessary here. Therefore, the selection of features increases efficiency and reduces the amount of time needed by users during the classification. ANOVA is also used to choose the right sound feature because it is a good way to find which feature is important for telling the difference between clearly different sound classes (Shetty, Patnaik and Prasad, 2022).

In this study, an ANOVA-based selection method is used to find how well different features in Kurdish dialects can tell them apart. The F-statistic operates within the selection process so scientists can determine feature significance, and lower p-values suggest better ranks. A p-value shows the chance of getting the F-statistic measurement or values higher than it when both assumptions of similar feature means across dialect groups are true. As p-values go down, the strength of the evidence against the null hypothesis goes up. This shows that the feature has big differences in mean values between classes, which makes it better at telling them apart.

All features undergo separate individual tests during the evaluation method. This evaluation scrutinizes each feature individually, allowing for the independent measurement of its discriminative power. The testing of features separately eliminates the requirement to define or evaluate groups of features, making the process simpler without compromising the discriminatory effectiveness. This kind of method lets you find features by evaluating each one separately, which leads to the choice of k = 200 features that produce the best dialect separation.

The data enhancement features take MFCC and Delta MFCC from the audio signals and combine them with Mel Spectrogram and Spectral Contrast. The computed features through Librosa represent audio signals in an extensive manner. The feature vectors from each sample are judged by ANOVA-based FS, which ranks the features based on their F-statistic values. The top-k (k = 200) features, which provide effective separation between Kurdish dialects, were selected through this process. The selection approach maintains a proper alignment between computational efficiency and performance outcomes.

VI. Experimental Results

The present section provides a detailed discussion of the performance assessment of the experimental results based on our proposed Kurdish dialect recognition model. As a result, we study the impact of DA and FS on the accuracy of our classifier and compare it with the performance of the model described in the previous section. Furthermore, we evaluate the proposed approach's accuracy for classification through confusion matrices. These empirical results clearly confirm the substantial improvements introduced in DA and appropriate FS. The proposed novel Kurdish Dialect Recognition Model is found to be very accurate and significantly outperforms the previous methods.

A. Dialect Recognition

Table II also provides the analysis of the present suggested model and the preceding exploration that we conducted recently. There is a detailed study of the sensitivity analysis of the environmental parameters such as pre- and post-augmentation signal length, DA methodologies, feature extraction methods, FS, and accuracy. From Table II, one can easily see that our proposed model performs significantly better than the current state-of-the-art model suggested in Ghafoor et al. (2021). Our model gives 99.42% accuracy against 95.5% of the existing model. White noise addition alongside time-stretching modifications enabled the dataset to improve through modifications of the initial speech signals. Time-stretching techniques lengthened the signals without modifying their basic features and white noise generated variations in the audio to enhance data augmentation. The model obtained access to a much broader dataset because of these enhancements which enabled it to gain knowledge from multiple input patterns leading to better dialect generalization. Due to this technique, the model was able to learn appropriately from a larger and more diverse dataset. It was achieved utilizing new features from MFCC, Mel Spectrogram, Poly-feature, and Contrast and the approach for FS by filter.

Fundamentally, confusion matrices are adopted to assess our proposed Kurdish dialect recognition model's classification efficiency and discuss its outcomes thoroughly. More comprehensive confusion matrices of each Kurdish dialect and other evaluation metrics such as Producer Accuracy (Precision), Recall, and the F1-score are shown in Table III below. The confusion matrices show that the proposed model achieved excellent classification results for each Kurdish dialect. Based on this study, the system can tell the difference between Badini, Hawrami, and Sorani dialects with very high scores for recall, F1-score, and producer accuracy (precision). Notably, the model can classify the Hawrami and Sorani dialects with perfect results in terms of the producer's accuracy.

The results from the experiments underscore the enhancement that has been achieved through DA, advanced methods within feature extraction, and selective FS. Our proposed Kurdish Dialect Recognition Model improves the previously successful approaches to show that the proposed model, with an accuracy of 99.42%, is extremely effective. This critical advancement in comparison to the prior work can demonstrate the potential and performance of the hinted model in these language identification tasks. The model's robust classification capabilities are further highlighted by

TABLE II
ACCURACY OF KURDISH DIALECT RECOGNITION MODELS

Methods	Length of the signal before DA	Use Data Augmentation	Length of the signal after DA.	Feature extracted	Feature selection	Accuracy (%)
(Ghafoor, et al., 2021)	895 samples	Null	895 samples	MFCC	Null	95.5
Proposed model	6000 samples	White Noise, time stretching	18000 samples	MFCC, Mel spectrogram, poly-feature, contrast	Filter-based	99.42

DA: Data augmentation

TABLE III
CONFUSION MATRICES WITH PRODUCER ACCURACY AND USER ACCURACY

Classes	TABLE	Hawrami	Sorani	Classification (support)	Producer accuracy (precision) (%)	Recall (%)	F1-score (%)
Badini	1190	5	2	1197	99	99	99
Hawrami	6	1219	7	1232	100	99	99
Sorani	0	1	1170	1171	99	100	100

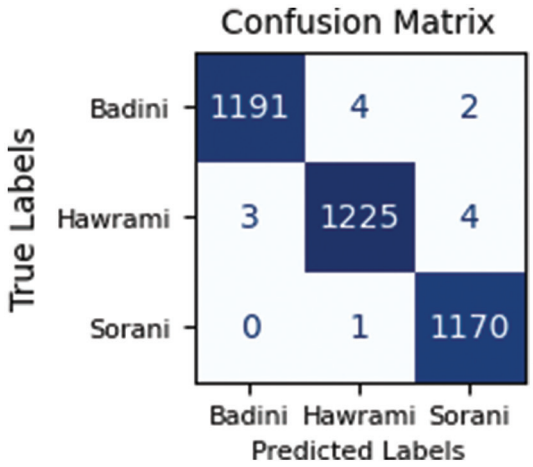

Fig. 6. Confusion matrix for Kurdish dialect classification.

analyzing the confusion matrices. The extensive literature review will allow understanding the strengths and weaknesses of the proposed model and its broad potential applications in the field of language recognition and classification.

The model's performance evaluation includes analysis through the confusion matrix in Fig. 6. The confusion matrix displays the total of accurate and inaccurate dialect predictions between Badini, Hawrami, and Sorani dialects. The numbers in the diagonal elements show correct dialect predictions, while the data points in off-diagonal elements indicate incorrect predictions.

B. Producer Accuracy and F1-Score

In our assessment of the Kurdish Dialect Recognition Model, two crucial performance metrics emerge: Producer Accuracy (Precision) and F1-Score. All things considered, these measures show how well the program identifies dialects and how accurate its classifications are.

Table III shows the Producer Accuracy (Precision) measure, which is the percentage of correct predictions for each dialect class out of all the cases that were put into that class. This metric is an important way to check if the model can get rid of false positives, which would mean high accuracy in the identified dialect. As observed in Fig. 7, the proposed model also exhibits excellent performance in terms of Producer Accuracy scoring between 99% and 100% in all the dialects. This accomplishment shows that our model correctly predicts a dialect, which in language recognition tasks is crucial because misclassification has real-world consequences. The results of Producer Accuracy show that the model is stable and confirm that it can be used to identify languages and other related sciences.

The F1 score is one of the most important features, especially when the combination of accuracy and both false-positive or false-negative results is needed. We measure the proposed model's performance for the dialect identification task using precision and recall. While every dialect is of utmost importance in language recognition scenarios, the F1-score offers a robust means of evaluating the model's performance without compromising precision and recall. With respect to performance, the F1 scores for our model are explicitly high, with percentages of 99 and 100% for every dialect. This achievement proves that the model has high precision and recall to reduce false positives and false negatives. The high F1 scores show that the model is good at telling the difference between the dialects we use in many language-related projects.

To sum up, the KDRM model is very good at identifying things and classifying them, with both high producer accuracy and high results classification effectiveness (a high F1 score). The aggregate view of the metrics provides strong evidence of the model's accuracy and usefulness in dialect identification. More importantly, their description is truly inspiring, as numerous approaches based on the model have found widespread application in various language recognition tasks.

VII. DISCUSSION

DA methods, which include noise addition and time-stretching operations, improve the diversity and resistance of training data datasets. By adding white noise to data simulation, the model becomes more adaptable to real-world environments, and TS varies speech speed to make the model more flexible with varied speakers and accents. The techniques function as effective data expansion methods in low-resource settings by requiring no new recorded data.

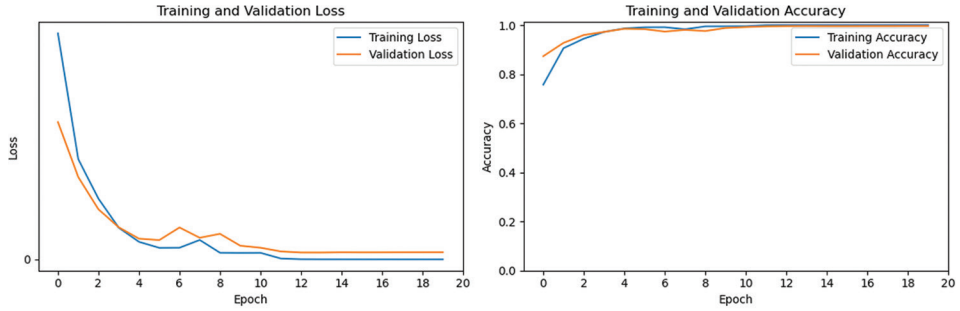

Fig. 7. Training and validation accuracy and training and validation loss.

The process of selecting features makes models more accurate while cutting down computational requirements so accuracy improves through better FS. The process enables better generalization with new data because it removes unneeded information to prevent overfitting.

It works well with both DA and FS technology because DA increases the variety of datasets and FS makes sure that datasets are ready for efficient learning. The joint use of DA techniques with FS methods creates powerful performance enhancement, which enables robust computational processing in Kurdish dialect classification methods.

The results of the experiments show that every part of the proposed approach makes model execution work better in important ways. A performance review of the model, which operates under multiple configurations, appears in Table IV. The integrated method of DA along with FS produces the best accuracy level.

Table V offers a more exhaustive cross-section of the features extracted and a more precise classification of several methods used in the field of signal processing (or a related field). We compare the approaches based on the feature sets they use, taking into account their respective accuracy levels. Each one of them incorporates a different approach to feature choice and to model training, which in its turn has an impact on the results.

The approach adopted from Al-Talabani, Abdul and Ameen (2017) involves local binary patterns (LBP), which is the premier method of feature extraction, and linear predictive coding (LPC). While LPC is used in speech processing for modeling the spectrum of a signal that is more or less constant over time, LBP is famous as a texture image descriptor. Moreover, even when the method is worked out using both of them at the same time, it only reaches a level of about 89.6%, which is somewhat lower than that of the more modern techniques. The findings show that while LBP and LPC can extract useful features for classification, they may not apply all the structures in the data to achieve optimal classification. Our group in prior work (Ghafoor, et al., 2021) used a 1D CNN for classification through MFCC feature set. MFCCs prove to show a high degree of efficiency in the signal processing of audio-type signals since they represent the power spectrum of the sound. The application of these coefficients in the 1D CNN structure improved the classification accuracy to a rate of 95.5%. This result validates the effectiveness of deep learning models, especially CNNs, in modeling time series data like the audio signals as against traditional feature extraction approaches.

TABLE IV
Impact of Data Augmentation and Feature Selection on Model Performance

Experiment setup	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Baseline model	96.5	96	96.3	96
With augmentation only	95.8	95.5	95.6	93.5
With feature selection only	94.2	94.0	94.1	93.0
Proposed method (Aug+Selection)	99.61	99.61	99.61	99.61

TABLE V
Comparison Proposed Model with Previous Research

Methods	Feature sets	Accuracy (%)
(Al-Talabani, Abdul and Ameen, 2017)	Local binary patterns+linear predictive coding	89.6
(Ghafoor, et al., 2021)	MFCC (Mel-frequency cepstral coefficients) in 1D CNN	95.5
(Karim, et al., 2024)	MFCC+Mel Spectrogram+Poly-feature+Contrast	96.5
Proposed model	Data Augmentation+ANOVA-based feature selection in 1D CNN	99.42

ANOVA: Analysis of variance, CNN: Convolutional neural network

Karim et al. (2024) extended this list of features by including MFCC + MelS + Poly-features + Contrast. Adding the Mel spectrogram introduces the time-frequency representation of the signal, which is particularly useful if temporal and spectral information is required. Poly-features refer to different signal characteristics, and contrast is added here to offer a comparison between the signal areas. This kind of work provides the so far largest feature set to achieve an accuracy of 96.5% in classifying the galaxies, which proves the great improvement against previous approaches. However, the strengthened interior combination still is not able to reach the pinnacle of the accuracy recorded in the current research.

Finally, this paper also presented a proposed model that comprises an advanced pipeline with DA and FS that employs ANOVA, as well as the 1D CNN. DA is accepted as a known method for increasing the training dataset, reducing overfitting, and increasing the model's ability to generalize. Another way to find the most expressive features in a classification problem is a technique known as ANOVA-based attribute ranking – it compares the differences

between various feature groups and how they contribute to a classification task. This approach enriches the features vector so the model contains only the most discriminating features. Using the aforementioned techniques, we were able to recognize the proposed study with 99.42% accuracy, leaving far behind other methods. Given a perfect combination of stable course development, the principal component selection, and deep learning techniques within a perfectly tuned network architecture, this high accuracy has very significant implications for real applications. As a result, the study shows that using better methods for extracting features leads to more accurate classification, especially when neural solid network designs are combined with techniques for improving data. The proposed model is superior to previous models by containing a deep learning aspect and stringent FS using ANOVA. Obtaining the final accuracy of 99.42%, we indicate that the suggested method can be successfully applied to practical tasks where accuracy is critical.

VIII. Conclusion

This paper presents a detailed analysis of our Kurdish Dialect Recognition Model, demonstrating its ability to perform language recognition and classification. Strong experimental findings demonstrate that our model determines major potential for language identification work. The investigation reveals the model's effective discrimination abilities between Kurdish dialects and establishes potential for its methods to identify languages beyond Kurdish dialects. The strong model performance stems from the integration between feature extraction and data augmentation and feature selection methods which establishes the model as a promising technology for field advancement. The benefits of our model are: Our model employs a rigorous procedure for feature extraction and selectively utilizes DA. We also introduced white noise and time-stretching techniques to the suggested dataset so as to improve the training data and increase the probability of the model learning and performing optimally on more than just a larger data set. We have added new characteristics such as MFCC, Mel Spectrogram, Poly-feature, and Contrast, which significantly enhance the model's ability to identify dialect differences. Moreover, an ANOVA-based filter-based FS technique has minimized the span of feature space while preserving relevant data. As a result, the model is now more accurate and effective, and it does a good job of detecting and differentiating dialects.

In real life, model accuracy and categorical classification are important performance indicators for language recognition tasks. The high values of Producer Accuracy and F1-Score suggest that this is possible. The proposed Kurdish Dialect Recognition Model's results show that feature extraction, data enhancement, and FS are all very important parts of being able to recognize languages. The accuracy and precision, as well as the balanced classification, show the efficiency of the solid and well-developed machine-learning methodology of the proposed model. This work has paved the way for further research and development in language identification and classification on top of enhancing language recognition capabilities. The role of accurate and flexible language recognition systems increases rapidly due to the free and significant growth of our world and the change that occurs in it. With the proposed model, the problems listed can be fixed, and the model can be used in other language-related fields that have not been studied yet. This will help keep an understanding of the variety of languages in a globalized world.

When highlighting the advantages of the presented model, it is essential to mention potential shortcomings, such as possible biases due to the assumptions made in the course of the study. Focusing on our approach in detail demonstrates a few areas of improvement that could arrive at different feature extraction techniques and elevate FS methodology. Furthermore, learning Kurdish dialects is a good step toward making the app work better in places with different cultures and languages, which includes making it more flexible. This approach proves that our model is beneficial as it allows for keeping languages diverse and developing intercultural communication. In addition, the limitations handled and the more comprehensive applications demonstrated herein support the advancement of language recognition technology, which creates a foundation for future advancements in language identification and categorization.

IX. Future Work

Our model achieves its strength by having an independent feature extraction process combined with selection mechanisms that operate without language restrictions. Using MFCC and a Mel Spectrogram along with statistical selection methods helps the model find speech features that are common across languages. Through DA approaches, the model demonstrates robust performance because these methods replicate pronunciation variability as well as accent and noise patterns, which affect speech-based tasks universally.

The approach requires an evaluation of language datasets, which will establish its capability for cross-language generalization. Model adaptation performance across different languages can be measured by training it on various linguistic datasets that evaluate its ability to learn universal phonetic components independent of language variation. Transfer learning methods should be applied to enhance the model by adapting it to underserved language domains through small, labeled datasets. This study focuses on the Kurdish language while extending its research approach toward different languages. Future researchers will work on conducting thorough assessments of this model, which will lead to developing adaptation methods for broad multilingual usage.

References

Abdul, Z.K., Al-Talabani, A., and Abdulrahman, A.O. 2016. A new feature extraction technique based on 1D local binary pattern for gear fault detection. *Shock and Vibration,* 2016, p. 8538165.

Abdullah, A.A., Abdulla, S.H., Toufiq, D.M., Maghdid, H.S., Rashid, T.A., Farho, P.F., & Asaad, A.T. 2024. NER-RoBERTa: Fine-Tuning RoBERTa for Named Entity Recognition (NER) within low-resource languages. *arXiv preprint*

arXiv:2412.15252.

Aguiar, R.L., Costa, Y.M., and Silla, C.N. 2018. Exploring data augmentation to improve music genre classification with convnets. In: *2018 International Joint Conference On Neural Networks (IJCNN)*. IEEE, United States, pp. 1-8.

Al-Onazi, B.B., Nauman, M.A., Jahangir, R., Malik, M.M., Alkhammash, E.H., and Elshewey, A.M. 2022. Transformer-based multilingual speech emotion recognition using data augmentation and feature fusion. *Applied Sciences,* 12, p. 9188.

Al-Talabani, A.K., Abdul, Z.K., and Ameen, A.A. 2017. Kurdish dialects and neighbor languages automatic recognition. *ARO-The Scientific Journal of Koya University,* 5, pp. 20-23.

Bahari, M.H., Dehak, N., Burget, L., Ali, A.M., and Glass, J. 2014. Non-negative factor analysis of gaussian mixture model weight adaptation for language and dialect recognition. *IEEE/ACM Transactions On Audio, Speech, and Language Processing,* 22, pp. 1117-1129.

Cheng, W.K., Khairuddin, I.M., Majeed, A.P.A., and Razman, M.A.M. 2020. The Classification of heart murmurs: The identification of significant time domain features. *Mekatronika: Journal of Intelligent Manufacturing and Mechatronics,* 2, pp. 36-43.

Damskägg, E.P., and Välimäki, V. 2017. Audio time stretching using fuzzy classification of spectral bins. *Applied Sciences,* 7, p. 1293.

Das, A., Guha, S., Singh, P.K., Ahmadian, A., Senu, N., and Sarkar, R. 2020. A hybrid meta-heuristic feature selection method for identification of Indian spoken languages from audio signals. *IEEE Access,* 8, pp. 181432-181449.

Das, P.P., Allayear, S.M., Amin, R., and Rahman, Z. Bangladeshi dialect recognition using Mel frequency cepstral coefficient, delta, delta-delta and Gaussian mixture model. In: *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI),* IEEE, United States, pp. 359-364.

Ghafoor, K.J., Rawf, K.M.H., Abdulrahman, A.O., and Taher, S.H. 2021. Kurdish dialect recognition using 1D CNN. *ARO-The Scientific Journal of Koya University,* 9, pp. 10-14.

Hama Rawf, K.M., Abdulrahman, A.O., and Mohammed, A.A. 2024. Improved recognition of Kurdish sign language using modified CNN. *Computers,* 13, p. 37.

Hu, H., Tan, T., and Qian, Y. Generative adversarial networks based data augmentation for noise robust speech recognition. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, United States, pp. 5044-5048.

Kanda, N., Takeda, R., and Obuchi, Y. Elastic spectral distortion for low resource speech recognition with deep neural networks. In: *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*. IEEE, United States, pp. 309-314.

Karim, S.H.T., Ghafoor, K.J., Abdulrahman, A.O., and Rawf, K.M.H. 2024. A Multi-feature fusion approach for dialect identification using 1D CNN. *JOIV: International Journal on Informatics Visualization,* 8, pp. 1246-1252.

Khamparia, A., Gupta, D., Nguyen, N.G., Khanna, A., Pandey, B., and Tiwari, P. 2019. Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access,* 7, pp.7717-7727.

Kupryjanow, A., and Czyzewski, A. 2012. A method of real-time non-uniform speech stretching. *E-Business and Telecommunications: International Joint Conference, ICETE 2011*, *Seville, Spain, July 18-21*. Revised Selected Papers. Springer, Germany, pp. 362-373.

Li, X., Zhang, W., Ding, Q., and Sun, J.Q. 2020. Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *Journal of Intelligent Manufacturing,* 31, pp.433-452.

Lounnas, K., Lichouri, M., and Abbas, M. 2022. Analysis of the effect of audio data augmentation techniques on phone digit recognition for algerian arabic dialect. *In: 2022 International Conference on Advanced Aspects of Software Engineering (ICAASE)*. IEEE, United States, pp. 1-5.

Ma, R., Tao, P., and Tang, H. 2019. Optimizing data augmentation for semantic segmentation on small-scale dataset. In: *Proceedings of the 2nd International Conference on Control and Computer Vision*, pp. 77-81.

Moreno-Barea, F.J., Jerez, J.M., and Franco, L. 2020. Improving classification accuracy using data augmentation on small data sets. *Expert Systems with Applications,* 161,

Mulahuwaish, A., Gyorick, K., Ghafoor, K.Z., Maghdid, H.S., and Rawat, D.B. 2020. Efficient classification model of web news documents using machine learning algorithms for accurate information. *Computers and Security*, 98, p. 102006.

Nguyen, T.S., Stueker, S., Niehues, J., and Waibel, A. 2020. Improving sequence-to-sequence speech recognition training with on-the-fly data augmentation. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, United States, pp. 7689-7693.

Nugroho, K., and Noersasongko, E. 2022. Enhanced Indonesian ethnic speaker recognition using data augmentation deep neural network. *Journal of King Saud University-Computer and Information Sciences,* 34, pp. 4375-4384.

Peddinti, V., Chen, G., Manohar, V., Ko, T., Povey, D., and Khudanpur, S. 2015. Jhu aspire system: Robust lvcsr with tdnns, ivector adaptation and rnn-lms. In: *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, United States, pp. 539-546.

Ragni, A., Knill, K.M., Rath, S.P., and Gales, M.J. Data augmentation for low resource languages. In: *Interspeech 2014: 15th Annual Conference of the International Speech Communication Association, 2014*. International Speech Communication Association (ISCA), pp. 810-814.

Rawf, K.M.H., Karim, S.H.T., Abdulrahman, A.O., and Ghafoor, K.J. 2024. Dataset for the recognition of Kurdish sound dialects. *Data in Brief,* 53, p. 1.

Rebai, I., Benayed, Y., Mahdi, W., and Lorré, J.P. 2017. Improving speech recognition using data augmentation and acoustic model fusion. *Procedia Computer Science,* 112, pp. 316-322.

Rituerto-González, E., Mínguez-Sánchez, A., Gallardo-Antolín, A., and Peláez-Moreno, C. 2019. Data augmentation for speaker identification under stress conditions to combat gender-based violence. *Applied Sciences,* 9, p. 2298.

Salamon, J., and Bello, J.P. 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters,* 24, pp. 279-283.

Sangwan, P., Deshwal, D., and Dahiya, N. (2021). Performance of a language identification system using hybrid features and ANN learning algorithms. *Applied Acoustics,* 175, p. 107815.

Shetty, N., Patnaik, L., and Prasad, N. 2022. Emerging research in computing, information, communication and applications proceedings of ERCICA 2022. In: *Proceedings of ERCICA*, p. 1.

Tu, Z., Liu, B., Zhao, W., Yan, R., and Zou, Y. 2023. A feature fusion model with data augmentation for speech emotion recognition. *Applied Sciences,* 13, p. 4124.

Tubishat, M., Abushariah, M.A., Idris, N., and Aljarah, I. 2019. Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. *Applied Intelligence,* 49, pp. 1688-1707.

Turab, M., Kumar, T., Bendechache, M., and Saber, T. 2022. Investigating multi-feature selection and ensembling for audio classification. *arXiv preprint arXiv:2206.07511*.

Wu, T., Duchateau, J., Martens, J.P., and Van Compernolle, D. 2010. Feature subset selection for improved native accent identification. *Speech Communication,* 52, pp. 83-98.

Zheng, Q., Yang, M., Tian, X., Jiang, N., and Wang, D. 2020. A full stage data augmentation method in deep convolutional neural network for natural image classification. *Discrete Dynamics in Nature and Society,* 2020, p. 4706576.

Zhou, H., Wang, X., and Zhu, R. 2022. Feature selection based on mutual information with correlation coefficient. *Applied Intelligence,* 52, pp. 5457-5474.