International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT22839

Spoken Language Recognization Based on Features and Classification Methods

Pooja Bam[1], Sheshang Degadwala[2], Rocky Upadhyay[3], Dhairya Vyas[4]

[1]Research Student, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

[2]Associate Professor, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

[3]Assistant Professor, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

[4]Managing Director, Shree Drashti Infotech LLP, Vadodara, Gujarat, India

Article Info

Volume 8, Issue 3
Page Number : 20-29

Publication Issue :
May-June-2022

Article History
Accepted: 01 May 2022
Published: 08 May 2022

ABSTRACT

In Western countries, speech-recognition applications are accepted. In East Asia, it isn't as common. The complexity of the language might be one of the main reasons for this latency. Furthermore, multilingual nations such as India must be considered in order to achieve language recognition (words and phrases) utilizing speech signals. In the last decade, experts have been clamoring for more study on speech. In the initial part of the pre-processing step, a pitch and audio feature extraction technique were used, followed by a deep learning classification method, to properly identify the spoken language. Various feature extraction approaches will be discussed in this review, along with their advantages and disadvantages. Purpose of this research is to Learn transfer learning approaches like Alexnet, VGGNet, and ResNet & CNN etc. using CNN model we got best accuracy for Language Recognition.

Keywords : Speech Recognition, Indian Language, Spoken Language, Pitch, Audio Feature, Machine Learning and Deep Learning

I. INTRODUCTION

Language identification refers to a machine's capacity to recognise spoken language. Detection of spoken language is done automatically using language recognition. Speeches are usually given by a stranger. The use of voice command systems to link humans and machines is becoming more common in today's world. Those who are fluent in the spoken language may now be confidently identified.

Consequently, persons in South Asian nations have been unable to fully benefit from advancements in

speech recognition technology since the development of speech detection algorithms for the Indic languages has been delayed as a result of this delay. Because of their complexity, Indian languages are difficult to convey on their own, and the multilingualism of these nations makes the task much more difficult. When speaking in this country, it is vital to identify the language of uttered words and phrases before attempting to recognise them since individuals seldom communicate in a combination of languages. Automated speech authentication is a technique for automatically distinguishing between different languages based on voice cues. This system can distinguish spoken segments and activate language-specific recognizers, which is useful in multilingual nations such as India, where speech recognition is difficult.

There are two primary approaches to speech recognition: acoustic and phonetic. Initially, acoustic approaches recover short-term speech spectrum features as a multidimensional vector. A statistical model is built for each language based on the extracted features. In acoustic-based SLR systems, Gaussian mixture models are the most often used model (GMM). The bulk of acoustic-based language recognition systems today employ the i-vector technique. It's the best in the field of language detection. This technique converts each speech file into a fixed-length vector. i-vectors are compressed speech signals used as input feature vectors in recognition systems' classification steps. Short-term acoustic characteristics are the easiest way to extract information from a speech stream. It is possible to extract higher-level speech information, such as phonetic information, from the voice signal. Phonetic-based SLR systems utilise the speech signal's phonetic information.

In the next paragraphs, we'll discuss each of the following topics: In Section II, we'll take a look at some important advancements in speech recognition. Here, in Section III, a full description of the many approaches employed to create this framework is provided. Section IV is dedicated to comparative studies and discussions. Finally, some suggestions for further research are made at the conclusion of this study.

II. Related Works

Bach Chu Paul et al. [1] Employed a pre-processing phase, followed by a pitch and Mel Frequency Cepstral Coefficients (MFCC) feature extraction approach, and lastly a Long Short-Term Memory, Deep Neural Network sequence classification method for accurate spoken language recognition. They employed the International Institute of Information Technology, Hyderabad (IIIT-H) Indic voice corpus in their proposed language identification effort, which included seven languages and 1000 spoken sentences for each language. As a result, our language identification model employed 7000 audio samples in total. With a huge dataset, the performance suffers marginally. The suggested technique performs well in terms of accurate language identification in a short amount of time.

Hung-Shin Lee et al., [2] the suggested technique obtained relative error rate reductions of 52 percent, 46 percent, 56 percent, and 27 percent when compared to the sequence-based PPR-LM, PPR-VSM, and PPR-IVEC methods, as well as the lattice-based PPR-LM method. They offer a novel learning technique for language verification and dialect/accent recognition that is based on subspace-based representation and can extract hidden phonotactic features from utterances. Subspace learning based on kernel machines, such as support vector machines and recently constructed subspace-based neural networks, is the subject of the second section (SNNs). They looked at temporal complexity in this work. They suggested a novel phonotactic representation of an utterance that is based on the idea of linear subspace and is neither equivalent nor reducible to a distributional or vectorial representation.

Musatafa Abbas et al., [3] Due to the random selection of weights inside the input hidden layer in the whole learning process of this model is not fully effective (i.e., optimised). The LID learning model used in this work is ELM, which is focused on extracting standard features. The generated results are based on LID with the same benchmarked data set derived from eight languages, which showed that the particle swarm optimisation–extreme learning machine LID (PSO–ELM LID) performed better than the ESA–ELM LID, with an accuracy of 98.75 percent compared to 96.25 percent for the ESA–ELM LID. For the objective of boosting the LID accuracy, the previous ELM-based learning model was extended as ESA–ELM. The study's next goal is to create a LID system that can conduct feature extraction and classification online while also including real-time components.

Himadri Mukherjee et.al [4] colleagues categorise spoken language using spectrograms (for image data) and deep learning, according to their research. In contrast to previous work, they propose to use voice signal patterns for spoken language detection in this study rather than image-based features, as has been done before. As a consequence, they recommend in this research that voice signal patterns be used for spoken language recognition rather than image-based features. The notion was sparked by the fact that speech transmissions may be interpreted and represented in certain situations. In all, five of the seven languages can be identified with 100 percent accuracy, with the other three languages detecting with more than 99 percent accuracy. Results were compared to those of earlier research, as well as common auditory and textural characteristics. The method provided correct results in all areas. They'll also make use of vocal activity detection to improve the overall efficiency of our system.

Shabnam Gholamdokht Firooz et al., [5] the process of phone identification is a time-consuming activity that may result in a considerable computational cost for the whole SLR system, particularly in its first phase. The conditional cascade structure is presented in this work in order to accelerate the combined system of phonetic and auditory approaches, which is currently being investigated. The final confidence ratings for each language are determined using a heuristic method based on the distribution of the acoustic score for each language. As a backup in the event that the proposed method's acoustic systems fail to properly identify the language of incoming speech with high confidence, a phonetic system is utilised, with the scores from both systems being pooled.

Dilip Singh Sisodia et al., [6] the most common difficulty encountered when converting spoken utterances into writing is correctly identifying the speaker's native language. The effectiveness of ensemble learning approaches for categorising spoken languages is investigated in this study. It is necessary to use significant features such as Mel–frequency cepstral coefficients (MFCC) in order to extract information from audio samples taken from the speaker. The features of audio signals known as the MFCC and DFCC are utilised to recognise spoken language. In this experiment, audio recordings of recorded conversations from five different languages are used to perform the experiments. Extensive usage of extra trees ensemble learners resulted in the identification of the best classifier among all ensemble models used for training and testing, with an accuracy of 85.00.

Gundeep Singh et al. [7] define spoken language identification as the act of detecting language from an audio recording made by an unknown speaker, regardless of gender, speaking style, or differentiating age differences (SLID). The model works with audio files, which it converts into spectrogram images using a spectrogram generator. An artificial convolutional neural network (CNN) is used to highlight essential characteristics or features that may be utilised to rapidly determine the result of the experiment. On various audio recordings, an experiment was carried out with the use of the Kaggle dataset named spoken

language identification. Two new contributions are made to the field of spoken language identification in this publication. This approach has a 98 percent accuracy rate and has yielded outstanding results in the field of medicine. Second, they perform language identification using the Bernoulli Naive Bayes approach on a dataset consisting of 22 languages. When comparing CNN and model fitting data, it takes a bit longer to complete the comparison.

Himanish Shekhar Das et al., [8] automatic language identification (LID) is a tough research topic in the realm of speech signal processing. It is used as the front end for many different applications, including multilingual conversational systems and spoken language translation. It is also used for spoken document retrieval and human-machine interaction. When it comes to automatic speech recognition (ASR) systems, language is often necessary. In most cases, humans are educated in the manner of a PPRLM model, which is why deep models that can imitate the PPRLM model are something to keep an eye on in the future.

Endah Safitri et.al [9] They use two different phonotactics techniques: Phone Recognition Followed by Language Modelling (PRLM) and Parallel Phone Recognition Followed by Language Modelling (PPRLM) are two methods of phone recognition followed by language modelling (PPRLM). Research on local languages in the field of spoken language identification (spoken LID) helps to extend the reach of technology to those who talk in their native language. When phone recognizers trained for English and Russian are used, the PRLM technique obtains the highest accuracy, with an average accuracy of 77.42 percent and 75.94 percent, respectively, for the English and Russian languages.

Panikos Heracleousetal [10] Deep learning (DL) is used to investigate spoken language identification, and the i-vector paradigm is provided. Deep neural networks (DNN) and convolutional neural networks (CNN) are used in this comparative study, which is provided (CNN). Previous research has shown that the DNN is capable of distinguishing between spoken languages. Utilizing the NIST 2015 i-vector Machine Learning Challenge challenge, the proposed techniques are examined for their effectiveness in the detection of 50 in-set languages. In addition, the data illustrate the applicability of CNN and i-vectors in recognising spoken languages, which is a promising development. When i-vectors are used, the results are positive, and they demonstrate that CNN and DNN are both capable of detecting spoken language well.

Radek Fer et al. in [11], The subject of spoken language recognition (SLR) is investigated by particular attention paid to the practical features of training bottleneck networks and their incorporation into SLR. Using qualities of monolingual and multilingual features, they show that multilingual training is appropriate for SLR training purposes. Using senones as targets, they illustrated and investigated the key downsides of expanding the output layer size, with a particular focus on the size of the input temporal context, in addition to the benefits. As reported in that paper, a baseline monolingual feature trained on a language that was similar to the test language performed better than multilingually trained features.

Mohit Dua, R et al. [12] construct and evaluate a continuous Hindi language speech recognition system that has been discriminatively trained. In order to train the automatic speech recognition (ASR) system, the system utilises maximum mutual information and minimum phone error discriminative techniques, as well as a variety of Gaussian mixtures, among other methods of learning the language. According to the results, improving the performance of the ASR system by employing RNN-based language modelling is beneficial. According to the findings of the investigations, discriminative training methodologies such as MMI and MPE outperform the typical MLE strategy for Hindi voice recognition. Demonstrate a significant improvement in performance.

Raymond W.M. Ng et. al [14] discuss In order to adapt it to various languages, the baseline tokenizer was trained on English conversational telephone voice data, and then utilised unsupervised cross lingual adaptation to adapt it to other languages. By using score fusion to merge SLR systems with changed tokenizers, it was possible to cut the minimum detection cost function (min DCF) by 7-18 percent when compared to baseline setups without modified tokenizers. With the use of cross-entropy (CE) and state-level minimum Bayes risk (sMBR) objective functions, they demonstrate how to do unsupervised cross lingual adaptation of DNN phoneme recognisers in spoken language recognition systems. After replacing the main DNN with any customised DNN in English SLR test trials, a performance drop in the phonotactic systems was seen in the phonotactic systems. Future study will focus on improving the quality of the updated DNN as well as the use of various adaptation approaches and system combinations, such as early fusion, in order to improve the overall performance.

Musatafa Abbas et al. [15] selected ELM as the learning model for LID in [15] because of the usual feature extraction employed in the study. The random generated weights contained in the input's hidden layers, on the other hand, render the model's learning process completely ineffective or unoptimized in its entirety. This study proposes a new optimised genetic algorithm (OGA) that employs three different selection criteria (roulette wheel, K-tournament, and random) to select the appropriate initial weights and biases for the ELM's input hidden layer, thereby reducing classification error and improving the ELM's overall performance for low-level information retrieval (LID). In order to optimise GA, three unique selection criteria for parent selection are used. GA was used in order to get the intended outcome.

Dhawale et.al [18] This study investigates a variety of text summarizing strategies that have been employed in previous research in this field. The basic goal of document text summarizing is to organize significant phrases and reduce the size of the document by removing superfluous sentences and less relevant facts.

III. Proposed Methodology

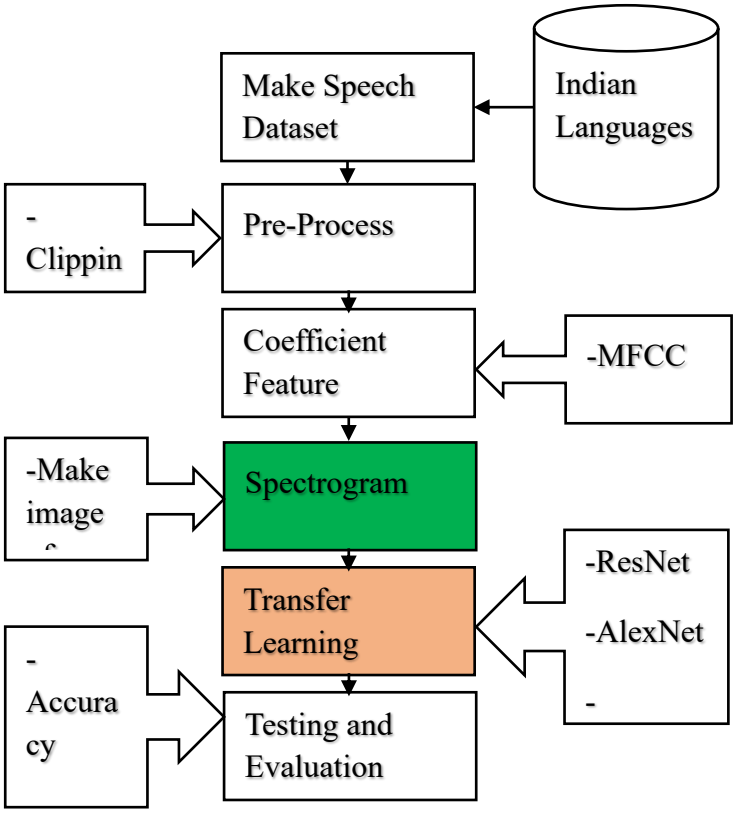

Fig. 1: Proposed System

We propose Transfer Learning based method for better classification. I have Used Different Types of Models. Using CNN Model Perform Batter among VGG, Alexnet and Resnet. Accuracy of Languages Recognition System is 99% in CNN Model.

Algorithm:

Step 1: Take input as .wav File
Step 2: Use noise removal approaches
Step 3: Extract Coefficient features
Step 4: Make Spectrogram
Step 5: Build Feature Map
Step 6: Train ResNet
Step 7: Classification and Evaluation

A. Datasets

The datasets in [1,4,6] are derived from the Indic speech corpus of the International Institute of Information Technology, Hyderabad (IIIT-H), which includes 1000 spoken sentences in each of seven different languages. Thus, they have utilized a total of 7000 audio samples in our language detection model.

There is a link to the data in [3] that may be found at (https://doi.org/10.6084/m9.figshare.6015173.v1).Additional data may be found on the author's website (http://www.ftsm.ukm.my/sabrina/resource.html), where it can be accessed.

The goal of this is to categories eight different languages, including English, Arabic, Malay, Spanish, French, German, Urdu, and Persian, into eight distinct categories. Each language has 15 utterances, with each speech lasting 30 seconds.

There are both target and non-target languages in the dataset set in [5]. With the use of this tool, SRL systems may be configured to identify Arabic, English, and Farsi languages by altering combination parameters or determining EER-based thresholds at operating locations across the system. Each target language has 200 files, while the non-target set has about 1000 files. The development files are typically about 30 seconds in duration.

The Kaggle dataset "spoken language identification" was used to analyses [7] distinct audio samples. These files include 10 second utterances, which are broken up into separate files.

B. Pre-Process [1,3,5,6]

- In the second step of audio processing, known as "clipping," audio signals are broken into frequency frames that are of the same size. Once this is done, use the windowing feature to remove the borders. An energy spectrum with no cross-overs at zero crossing rate. Remove background noise and unspoken information from an audio clip. Listeners should expect to hear 30 seconds of emotional variance in each track. Based on this evaluation, the start and finish points are determined.
- Additional refinement of log-Mel spectrograms may be achieved by the removal of background noise from audio recordings. You may enhance the data by utilizing numerous techniques, such as pitch shifting and cropping and rotating and flipping as well as adding random noise and adjusting the audio speed. This aids in making neural networks more resilient to changes that may occur in real-world circumstances.

C. Feature Extraction

- The pitch and 14 Mel Frequency Cepstral Coefficients (MFCC) were calculated as our feature for the voiced activity zone of the said phrase by chopping the signal into short chunks of frames. The signal is shortened into a 25ms segment with 50% overlap with the previous segment termed a frame, rather than detecting the characteristics of the whole phrase. For this reason, overlapping is employed because of the quick changes in voice signal and the fact that some linguistic information is transferred into the next frames. Each frame comprises 400 samples, or 80 frames per second, making it possible to shoot at a rate of one frame per second.
- A spectrogram is a visual depiction of the strength, or sound, of a signal over a long period of time utilizing different frequencies within a specific waveform structure in waveform analysis. The graph also shows how energy levels fluctuate over time.

D. Classification Methods

TABLE I. CLASSIFIER METHODS

ResNet [13,16,17]	It is possible to skip connections. It makes use of batch normalization to boost efficiency while maintaining accuracy.	Implementation is time-consuming.
AlexNet [17]	Unlike a convolutional layer, which depends on local spatial coherence and a narrow receiving field, a fully connected layer learns features from all of the combinations of the features of the preceding layer.	Complicated layers with many connections are very computationally costly to create.
VGGNet [11,9]	It only contains 80 percent of the whole number of parameters.	Accuracy decreases in a very progressive manner.
Inception V3 [17]	Allows for any conceivable combination of layers to be used.	Intensive training is required. Cost of calculation is really high.

IV. Result and Analysis

1) Loading Images: -

```
loading images...
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 127, 127, 64)      832

 max_pooling2d (MaxPooling2D  (None, 63, 63, 64)       0
 )

 dropout (Dropout)           (None, 63, 63, 64)        0

 conv2d_1 (Conv2D)           (None, 62, 62, 32)        8224

 max_pooling2d_1 (MaxPooling  (None, 31, 31, 32)       0
 2D)

 dropout_1 (Dropout)         (None, 31, 31, 32)        0

 flatten (Flatten)           (None, 30752)             0

 dense (Dense)               (None, 128)               3936384

 dropout_2 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 7)                 903

=================================================================
Total params: 3,946,343
Trainable params: 3,946,343
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2: CNN Model Architecture

2) Compile Model For Traning: -

```
#Compile the model
model.compile(loss='categorical_crossentropy',
              optimizer='adam',
              metrics=['accuracy'])
#Fit the model
history = model.fit(X_train,y_train, epochs=10,batch_size=128,verbose=1)
```

```
Epoch 1/10
50/50 [==============================] - 18s 166ms/step - loss: 2.1165 - accuracy: 0.1805
Epoch 2/10
50/50 [==============================] - 8s 162ms/step - loss: 1.5142 - accuracy: 0.4149
Epoch 3/10
50/50 [==============================] - 8s 163ms/step - loss: 0.9399 - accuracy: 0.6594
Epoch 4/10
50/50 [==============================] - 8s 163ms/step - loss: 0.5828 - accuracy: 0.7902
Epoch 5/10
50/50 [==============================] - 8s 163ms/step - loss: 0.3987 - accuracy: 0.8600
Epoch 6/10
50/50 [==============================] - 8s 163ms/step - loss: 0.3214 - accuracy: 0.8883
Epoch 7/10
50/50 [==============================] - 8s 163ms/step - loss: 0.2744 - accuracy: 0.8954
Epoch 8/10
50/50 [==============================] - 8s 164ms/step - loss: 0.2284 - accuracy: 0.9170
Epoch 9/10
50/50 [==============================] - 8s 164ms/step - loss: 0.2279 - accuracy: 0.9121
Epoch 10/10
50/50 [==============================] - 8s 162ms/step - loss: 0.1775 - accuracy: 0.9340
```

Fig. 3: CNN Model Train

3) Confusion Matrix :-

```
Confusion Matrix
[[ 86   0   0   0   3   0   0]
 [  0  93   0   0   0   0   0]
 [  0   0 100   0   0   1   0]
 [  0   0   0  94   0   0   0]
 [  0   0   0   0 104   0   1]
 [  0   0   0   3   0 110   0]
 [  0   0   0   0   0   0 105]]
Classification Report
              precision    recall  f1-score   support

     Bengali       1.00      0.97      0.98        89
       Hindi       1.00      1.00      1.00        93
     Kannada       1.00      0.99      1.00       101
   Malayalam       0.97      1.00      0.98        94
     Marathi       0.97      0.99      0.98       105
       Tamil       0.99      0.97      0.98       113
       Telgu       0.99      1.00      1.00       105

    accuracy                           0.99       700
   macro avg       0.99      0.99      0.99       700
weighted avg       0.99      0.99      0.99       700
```

Fig. 4: CNN Model Parameters Calculation

4) Graph of Model Accuracy And Loss Plot: -

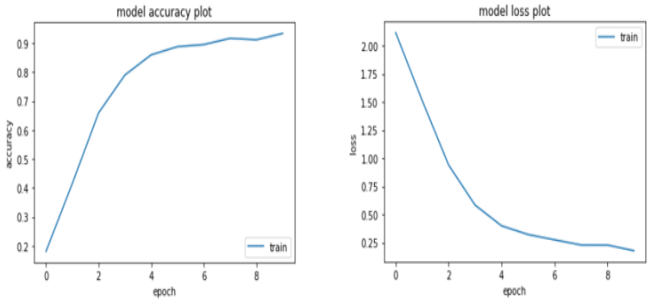

Fig. 5: CNN Model Accuracy and Loss Plot

5) Final Prediction :-

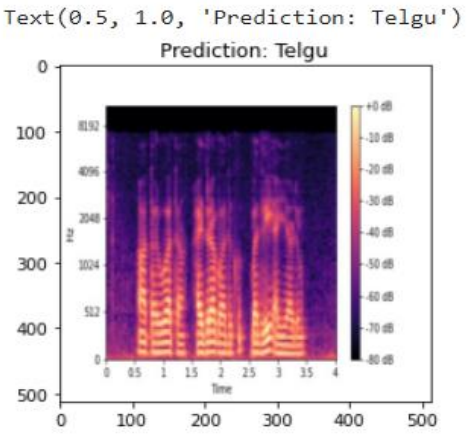

Fig. 6: CNN Model Test Result

TABLE II. ANALYSIS

Transfer Learning Model	Precision	Recall	Accuracy	F1-Score
CNN	99%	99%	99%	99%
VGG16	96%	96%	95%	96%
ResNet50	62%	59%	60%	56%
AlexNet	95%	93%	93%	93%

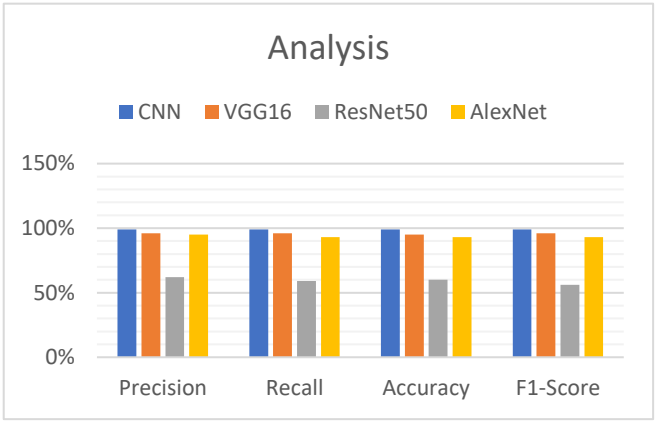

Fig. 7: CNN Analysis Graph

V. Conclusion

According to the review, several characteristics are used, but the MFFC feature is most often used in spoken language recognition systems. It was found that most writers utilize DNNs for classification, however other authors employ DL techniques, but they won't work with large feature vectors. Transfer learning algorithms, on the other hand, are popular right now and provide excellent results across a wide range of datasets. Several of them are discussed in this study, so that spectrogram feature with transfer learning techniques may deliver greater accuracy in less time calculation in the future. Using Transfer Learning Using CNN Model And Other Different

Types Of Model in Deep Learning But we got good accuracy in CNN it was 99%.

VI. References

1) B. Paul, S. Phadikar, and S. Bera, "Identification Using Deep Learning Approach," pp. 263–274.

2) H. S. Lee, Y. Tsao, S. K. Jeng, and H. M. Wang, "Subspace-Based Representation and Learning for Phonotactic Spoken Language Recognition," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 28, pp. 3065–3079, 2020, doi: 10.1109/TASLP.2020.3037457.

3) M. A. A. Albadr and S. Tiun, "Spoken Language Identification Based on Particle Swarm Optimisation–Extreme Learning Machine Approach," Circuits, Syst. Signal Process., vol. 39, no. 9, pp. 4596–4622, 2020, doi: 10.1007/s00034-020-01388-9.

4) H. Mukherjee et al., "Deep learning for spoken language identification: Can we visualize speech signal patterns?," Neural Comput. Appl., vol. 31, no. 12, pp. 8483–8501, 2019, doi: 10.1007/s00521-019-04468-3.

5) S. Gholamdokht Firooz, S. Reza, and Y. Shekofteh, "Spoken language recognition using a new conditional cascade method to combine acoustic and phonetic results," Int. J. Speech Technol., vol. 21, no. 3, pp. 649–657, 2018, doi: 10.1007/s10772-018-9526-5.

6) D. S. Sisodia, S. Nikhil, G. S. Kiran, and P. Sathvik, "Ensemble learners for identification of spoken languages using mel frequency cepstral coefficients," 2nd Int. Conf. Data, Eng. Appl. IDEA 2020, 2020, doi: 10.1109/IDEA49133.2020.9170720.

7) G. Singh, S. Sharma, V. Kumar, M. Kaur, M. Baz, and M. Masud, "Spoken Language Identification Using Deep Learning," Comput. Intell. Neurosci., vol. 2021, 2021, doi: 10.1155/2021/5123671.

8) H. S. Das and P. Roy, A deep dive into deep learning techniques for solving spoken language identification problems. Elsevier Inc., 2019.

9) N. E. Safitri, A. Zahra, and M. Adriani, "Spoken Language Identification with Phonotactics Methods on Minangkabau, Sundanese, and Javanese Languages," Procedia Comput. Sci., vol. 81, no. May, pp. 182–187, 2016, doi: 10.1016/j.procs.2016.04.047.

10) P. Heracleous, K. Takai, K. Yasuda, Y. Mohammad, and A. Yoneyama, "Comparative study on spoken language identification based on deep learning," Eur. Signal Process. Conf., vol. 2018-September, pp. 2265–2269, 2018, doi: 10.23919/EUSIPCO.2018.8553347.

11) R. Fér, P. Matějka, F. Grézl, O. Plchot, K. Veselý, and J. H. Černocký, "Multilingually trained bottleneck features in spoken language recognition," Comput. Speech Lang., vol. 46, pp. 252–267, 2017, doi: 10.1016/j.csl.2017.06.008.

12) M. Dua, R. K. Aggarwal, and M. Biswas, "Discriminatively trained continuous Hindi speech recognition system using interpolated recurrent neural network language modeling," Neural Comput. Appl., vol. 31, no. 10, pp. 6747–6755, 2019, doi: 10.1007/s00521-018-3499-9.

13) O. Giwa and M. H. Davel, "The effect of language identification accuracy on speech recognition accuracy of proper names," 2017 Pattern Recognit. Assoc. South Africa Robot. Mechatronics Int. Conf. PRASA-RobMech 2017, vol. 2018-January, pp. 187–192, 2017, doi: 10.1109/RoboMech.2017.8261145.

14) R. W. M. Ng, M. Nicolao, and T. Hain, "Unsupervised crosslingual adaptation of

tokenisers for spoken language recognition," Comput. Speech Lang., vol. 46, pp. 327–342, 2017, doi: 10.1016/j.csl.2017.05.002.

15) M. A. A. Albadr, S. Tiun, M. Ayob, and F. T. AL-Dhief, "Spoken language identification based on optimised genetic algorithm–extreme learning machine approach," Int. J. Speech Technol., vol. 22, no. 3, pp. 711–727, 2019, doi: 10.1007/s10772-019-09621-w.

16) Y. Ma, R. Xiao, and H. T. B, "An Event-Driven Computational System," vol. 1, pp. 453–461, 2017, doi: 10.1007/978-3-319-70136-3.

17) P. Beckmann, M. Kegler, H. Saltini, and M. Cernak, "Speech-VGG: A deep feature extractor for speech processing," no. May 2020, 2019, [Online]. Available: http://arxiv.org/abs/1910.09909.

18) Dhawale, Apurva D., Sonali B. Kulkarni, and Vaishali M. Kumbhakarna. "A Survey of Distinctive Prominence of Automatic Text Summarization Techniques Using Natural Language Processing." In International Conference on Mobile Computing and Sustainable Informatics, pp. 543-549. Springer, Cham, 2020.

Cite this article as :

Pooja Bam, Sheshang Degadwala, Rocky Upadhyay, Dhairya Vyas, "Spoken Language Recognization Based on Features and Classification Methods", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 3, pp. 20-29, May-June 2022. Available at doi : https://doi.org/10.32628/CSEIT22839
Journal URL : https://ijsrcseit.com/CSEIT22839