International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-6

An Intelligent Deep Learning Framework for Spoken Language Identification

K. Rajkumar[1], A. Swarshitha[2]
[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.
[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India

***Abstract* – Modern speech processing applications rely heavily on spoken language identification. This includes intelligent translation systems, voice-based human-computer interaction, and multilingual speech recognition. Communication technologies are enhanced by accurate voice recognition, which allows for efficient processing of multilingual speech data. The five most common Indian languages—Hindi, Bengali, Tamil, English, and Gujarati—are identified using a framework based on deep learning in this study. To classify languages, the suggested technique feeds a Deep Neural Network (DNN) with audio data extracted from speech signals using Mel-Frequency Cepstral Coefficients (MFCCs). To enhance classification performance, audio samples are preprocessed, features are extracted, and models are trained using publically accessible speech datasets. Common performance measures, including as F1-score, recall, accuracy, and precision, are used to assess the efficacy of the suggested model. The results of the experiments show that the deep learning model successfully differentiates between the target languages and shows impressive generalisation when applied to new audio data. In addition, the published web application is built using Django, which incorporates the trained model. Users may access the results of language recognition in real-time by uploading voice recordings and interacting with the interface. With its extensible design, the suggested framework may accommodate more languages and practical speech processing applications while simultaneously providing an efficient, scalable, and workable solution for multilingual spoken language recognition.**

***Keywords* – Spoken Language Identification, Deep Learning, Deep Neural Network, Mel-Frequency Cepstral Coefficients, MFCC, Speech Processing, Audio Classification, Django, Multilingual Recognition, Natural Language Processing.**

I. INTRODUCTION

The ability for intelligent systems to autonomously identify the language spoken in an audio stream has propelled Spoken Language Identification (SLI) to the forefront of speech processing and Natural Language Processing (NLP) research. An accurate spoken language recognition system is in high demand due to the fast development of speech-based communication technologies, Artificial Intelligence (AI), and Deep Learning (DL). Multilingual automated speech recognition, speech-to-speech translation, voice-controlled virtual assistants, intelligent surveillance, intelligent contact center automation, voice authentication, and human-computer interface are just a few of the many real-world applications that rely on these technologies. Improving the efficacy of multilingual communication systems and facilitating seamless engagement across varied linguistic settings are both made possible by reliably identifying the spoken language before further speech processing [1, 3].

Hundreds of languages and regional dialects are spoken in various parts of India, making it one of the most linguistically varied nations in the world. While a more diverse language pool certainly improves communication, it does present formidable obstacles to the advancement of automatic language recognition systems. Language categorisation is sometimes a challenging process due to variations in speaker features, recording setting, background noise, dialect, speaking tempo, accent, and pronunciation. As a result, developing an AI model that can reliably identify a variety of Indian languages has emerged as a major area of study in the field of speech analytics [4, 11].

In the past, systems for identifying spoken languages mostly used statistical learning methods for language classification, including GMMs, HMMs, and i-vector-based approaches. The approaches relied heavily on acoustic features that were hand-crafted and required a high level of domain knowledge for feature engineering, but they were successful in controlled environments. When dealing with complicated multilingual speech datasets that include differences in recording settings and speaker characteristics, their performance often deteriorated [5], [9].

Models can now automatically learn very discriminative representations from voice signals, thanks to recent advancements in Deep Learning, which has revolutionised spoken language detection. Modelling the nonlinear correlations in voice data successfully has led to Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks outperforming traditional machine learning methods. Improved generalisability to unknown speech samples and higher classification accuracy are two benefits of these designs, which lessen the need for human feature engineering [6, 12, 18].

One of the most popular representations for speech analysis is Mel-Frequency Cepstral Coefficients (MFCCs), which is one of many ways for extracting speech features. MFCC

features successfully extract useful spectrum information from voice signals while closely mimicking the perceptual properties of the human auditory system. Deep learning models are well-suited for spoken language recognition tasks because to these characteristics, which greatly enhance their capacity to differentiate between languages based on small auditory changes [11], [13].

This study proposes a framework for detecting the five most common Indian languages: Hindi, Bengali, Tamil, English, and Gujarati, using a Deep Neural Network (DNN). We guarantee the availability of different voice samples by collecting the speech recordings from publicly accessible datasets, such as Mozilla Common Voice and OpenSLR. Dataset preparation, audio format standardisation, duration normalisation, MFCC feature extraction, and label encoding are some of the pretreatment procedures that the acquired audio files go through before the model building process begins. By standardising the data, these preprocessing steps make it easier for the deep learning model to effectively learn audio patterns that are typical of the target language.

To enhance learning stability and classification performance, the suggested Deep Neural Network incorporates batch normalisation into its several fully linked dense layers. The model learns discriminative representations for each language category during training by being fed the extracted MFCC features as input to the network. Accuracy, Precision, Recall, and F1-Score are some of the common performance measures used to evaluate the trained model's efficacy in spoken language categorisation. The results of the experimental assessment show that the suggested framework maintains a high level of generalisability on new voice recordings and accurately recognises the chosen Indian languages.

The trained deep learning model is incorporated into a deployed web application based on Django to increase the practical usability of the suggested solution. Users may get real-time language recognition results via the web app's interactive platform, which allows them to contribute voice recordings. The frontend provides an easy-to-navigate interface for viewing prediction results, while the Django backend handles audio preprocessing, feature extraction, prediction creation, and database operations quickly. Applications in education, industry, and multilingual communication may all benefit from this deployment, which turns the learned deep learning model into a real intelligent system.

II. LITERATURE SURVEY

With its increasing relevance in AI, virtual assistants, multilingual speech processing, and intelligent communication systems, Spoken Language Identification (SLI) has emerged as a vibrant field of study. The fundamental goal of spoken language identification is to detect the language being said in an audio stream automatically prior to carrying out more advanced speech processing operations. Improved language classification accuracy under different acoustic settings has been the goal of several statistical, machine learning, and deep learning strategies suggested by researchers over the last two decades [1], [3].

The majority of the early language recognition systems that were based on statistical methods for extracting speaker-independent language features used GMMs, HMMs, and i-vector representations. For small datasets, these methods worked well, but they needed a lot of preprocessing and hand-crafted feature engineering. Their efficacy was often diminished when confronted with multilingual datasets that included varying degrees of accent, recording conditions, and prosody [5, 9].

By allowing automated feature learning straight from voice signals, the fast development of deep learning has drastically altered spoken language detection. One of the most popular designs is the Deep Neural Network (DNN), which can represent complicated nonlinear interactions among speech characteristics. By learning abstract acoustic representations with little to no dependence on human-designed characteristics, DNN-based systems have been shown to outperform conventional statistical approaches [3, 10].
A number of research projects have looked at the use of CNNs and LSTM networks for the classification of spoken languages. Local spectrum patterns in speech representations can be efficiently captured by CNN-based models, while long short-term memory (LSTM) networks can learn the long-term temporal relationships within speech sequences. By processing sequential audio data effectively, these designs have improved multilingual language detection tasks significantly [4, 8, 16].

When it comes to spoken language detection systems, feature extraction is king. Due to its near resemblance to the human auditory system, Mel-Frequency Cepstral Coefficients (MFCCs) have surpassed all other acoustic features in terms of their prevalence in speech representations. By enhancing the capacity of deep learning models to learn, MFCC features efficiently record data in the frequency domain while decreasing the amount of unnecessary audio components. The accuracy of language classifications is much improved when MFCC characteristics are combined with deep neural networks, according to many studies [11, 12, 18].

Data augmentation, end-to-end speech recognition models, pretrained neural networks, and transfer learning have all been the subject of recent research aimed at boosting language identification ability. Increasing dataset variety and decreasing overfitting are two ways these techniques increase model resilience, which is particularly useful when training data is limited. Also, in comparison to traditional recurrent networks, attention-based designs and transformer models are better at capturing contextual speech information, which has led to their recent surge in popularity [2, 14, 17].

Spoken language recognition has come a long way, but it's still not easy to identify languages that are linguistically close. Classification uncertainty arises in real-world circumstances when languages like Bengali and Hindi have similar phonetic traits. As a result, there is still a need for deep learning models that can develop acoustic representations that discriminate well while still performing well when applied to different speech datasets.

This paper presents a system for spoken language identification using Deep Neural Networks and Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction, driven by these issues. Hindi, Bengali, Tamil, English, and Gujarati are the five most widely spoken languages in India according to the suggested model. Along with that, the trained model is included into a web application that is deployed using Django. This application allows users to engage with an interactive interface to upload voice recordings and do real-time language detection. For applications that handle speech in more than one language, this framework offers a scalable and efficient solution.

Table I. Summary of Existing Spoken Language Identification Studies

Reference	Technique Used	Major Contribution
Previous Study 1	Gaussian Mixture Model (GMM)	Statistical language identification using acoustic features.
Previous Study 2	Deep Neural Network (DNN)	Improved language recognition through automatic feature learning.
Previous Study 3	CNN and LSTM	Enhanced multilingual speech classification using deep learning.
Previous Study 4	MFCC with Deep Learning	Improved classification accuracy using perceptual speech features.
Proposed Work	MFCC + Deep Neural Network	Accurate identification of five Indian spoken languages with Django-based web deployment

III. PROPOSED SYSTEM

The proposed spoken language identification framework set out to use Deep Learning techniques for the accurate categorisation of five of India's most popular languages. The whole procedure consists of collecting audio data, cleaning it, extracting MFCC characteristics, creating a model using a Deep Neural Network (DNN), testing it, and ultimately distributing it via a web app built using Django. In order to improve the suggested system's ability to classify languages, every detail is carefully considered.

The voice dataset was assembled using publicly available resources, including OpenSLR and the Mozilla Common Voice dataset. Languages used for the study include English, Bengali, Tamil, and Gujarati in addition to Hindi and Hindi. We begin by amassing around one thousand audio recordings for every language, and then we choose a statistically valid sample to ensure equitable training and evaluation. We collected audio samples from a wide range of speakers with different pronunciation patterns to assist the model learn the unique acoustic characteristics of each language.

To guarantee consistency in the dataset, many preprocessing steps are performed before training the model. To simplify processing, audio recordings that are available in numerous forms are converted using a common format. Due to the large discrepancy in spoken word recording durations, it is essential to trim lengthier recordings of superfluous portions and add silence to shorter ones in order to make all of the audio samples the same length. This preprocessing phase ensures that all speech samples are ready for feature extraction with an equal quantity of data.

After the audio files have been preprocessed, the Mel-Frequency Cepstral Coefficients (MFCCs) may be extracted. Because they mimic the human auditory perception system so well, MFCCs are able to faithfully represent the spectrum characteristics of speech, which has led to their extensive usage in language recognition and speech applications. We numerically encode the language labels and link them to the returned MFCC feature vectors to facilitate supervised learning. The trained and validation sets each get 80% of the prepared dataset, which allows us to evaluate the model's ability to learn and generalise.

In order to identify spoken languages, Deep Neural Networks (DNNs) are the primary classification models employed. Following a number of tightly coupled hidden layers, the network uses Batch Normalisation layers to improve training stability and convergence speed. To determine the likelihood of each target language, the output layer employs the Softmax activation function. In contrast, the hidden layers capture nonlinear links across MFCC features using Rectified Linear Unit (ReLU) activation. Using Sparse Categorical Cross-Entropy as the loss function, the Adam optimisation strategy is used to train the model for efficient multiclass language classification.

To ensure the predicted framework can provide correct predictions, we subject it to tests using unseen speech recordings after model training. Accuracy, Precision, Recall, and F1-Score are some of the common performance metrics used to assess the categorisation performance for each language category. When the trained model is integrated into a web application created using the Django framework, users may upload audio recordings of themselves speaking and get immediate results for spoken language recognition using an intuitive online interface.

International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-6

Table II. Proposed Framework Components

Stage	Description
Data Collection	Speech recordings collected from Mozilla Common Voice and OpenSLR datasets.
Data Preprocessing	Audio format conversion, duration normalization, and label encoding.
Feature Extraction	Mel-Frequency Cepstral Coefficients (MFCCs) extracted from speech signals.
Model Development	Deep Neural Network with Dense Layers, Batch Normalization, ReLU, and Softmax.
Performance Evaluation	Accuracy, Precision, Recall, and F1-Score evaluation.
Web Deployment	Django-based deployed web application for real-time spoken language identification.

IV. EXPERIMENTAL ANALYSIS

The efficiency of the proposed Deep Neural Network model in detecting five main Indian spoken languages is measured experimentally. All of the acquired audio files are preprocessed before training so that they are consistent in length, format, and feature representation. The MFCCs that were extracted are fed into the Deep Neural Network as features, and the language labels that correspond to them are employed as target classes. In an 80:20 split, the prepared dataset is split into training and validation subsets, so the model may learn from one part of the data while being tested on unknown samples, ensuring a trustworthy assessment.

To enhance learning stability and speed up convergence during training, the suggested model is built utilising Batch Normalisation in conjunction with several densely linked hidden layers. In order to capture nonlinear interactions among MFCC variables, the hidden layers utilise ReLU activation functions. In the output layer, the voice recordings are classified into one of the five target languages using the Softmax activation function. In order to update the model parameters efficiently, the Adam optimiser is used. As for multiclass classification, the loss function is Sparse Categorical Cross-Entropy.

We keep an eye on the Deep Neural Network's training and validation performance as it goes through 150 epochs of training. As the model learns significant language-specific acoustic patterns, the validation accuracy continually increases and the training loss lowers progressively. The model shows good generalisability without major overfitting, as seen by the tight correlation between training and validation performance.

Once the training phase is complete, unseen voice recordings are used to assess the proposed model's robustness. Using the same preprocessing and MFCC feature extraction algorithms used during training, a separate testing dataset is constructed using speech samples from English, Hindi, Bengali, Tamil, and Gujarati. This makes sure that the model's capacity to categorise new speech signals, which were not encountered when developing the model, is fairly evaluated.

In terms of classification performance, the testing findings show that the suggested Deep Neural Network can accurately identify the chosen spoken languages. Classification results and the confusion matrix show that Tamil, Gujarati, and Hindi are all correctly categorised, however English and Bengali are sometimes mixed up due to shared acoustic features in certain audio samples. In general, the suggested framework accomplishes about 90% accuracy in classification, proving that it is successful for identifying spoken languages in several languages.

The experimental results show that a powerful method for identifying spoken languages may be achieved by combining Deep Neural Network classification with MFCC-based feature extraction. In addition, the suggested framework is well-suited for practical applications in multilingual voice processing, as it allows for real-time language prediction via an easy online interface when the trained model is integrated into a Django-based deployed web application.

V. RESULTS AND DISCUSSION

English, Hindi, Bengali, Tamil, and Gujarati are the five spoken languages in India that make up the multilingual speech dataset that is used to assess the performance of the suggested Deep Neural Network. The model's classification performance improves steadily over the training process with increasing training epochs. The gradual reduction in training and validation loss, together with the corresponding increase in classification accuracy, indicates that the proposed model successfully learns meaningful language-specific acoustic patterns while maintaining good generalization capability on unseen speech data.

In order to assess the trained model's resilience, the testing phase employs a distinct set of audio recordings that have never been seen before. All testing samples undergo the same preprocessing operations, including audio standardization and MFCC feature extraction, before being supplied to the Deep Neural Network. In order to make sure that the results show how well the proposed framework predicts under real-world conditions, we use a consistent evaluation process.

The suggested model is effective in multilingual spoken language identification, as shown by experimental evaluation, which reaches an overall classification accuracy of about 90%. While there are only a handful of misclassifications between Bengali and English, the classification report shows that Gujarati, Hindi, and Tamil are recognised with high recall and precision values. The main reason for this behaviour is that these languages share a lot of acoustic traits and pronunciation patterns. The suggested Deep Neural Network reliably and effectively

distinguishes the selected Indian languages, according to the overall classification performance.

A comparative analysis with previously reported spoken language identification approaches indicates that the proposed Deep Neural Network provides competitive performance while utilizing MFCC-based speech features. The combination of multiple dense layers, batch normalization, and MFCC feature extraction enables the model to learn discriminative representations without requiring complex handcrafted feature engineering. The obtained results demonstrate that the proposed framework effectively balances classification accuracy and computational efficiency for practical language identification tasks.

Furthermore, the trained Deep Neural Network is deployed as a Django-based web application, allowing users to upload speech recordings and obtain real-time language identification results through a browser-based interface. The deployment of the trained model transforms the research framework into a practical intelligent system capable of supporting multilingual speech processing applications, educational platforms, voice-enabled services, and automated communication systems. The overall experimental findings confirm that the proposed methodology provides an efficient, scalable, and reliable solution for spoken language identification using deep learning techniques.

VI. WEB APPLICATION IMPLEMENTATION

The trained Deep Neural Network model is incorporated into a Django-based deployed web application to illustrate the practical usability of the suggested spoken language detection framework. You may get real-time language recognition results by uploading voice recordings to the web app's user-friendly platform. Users may efficiently access multilingual voice recognition with the suggested system's combination of deep learning and a web-based deployment environment. This is achieved without the need for direct user interaction with the underlying machine learning model.

Uploading audio, preprocessing it, extracting features, inferring a model, and generating results are all handled by the Django framework. When a user submits a voice recording using the web interface, the backend checks the file for errors, does any preprocessing that is needed, and then uses the same method as when training the model to extract Mel-Frequency Cepstral Coefficients (MFCCs). In order for the trained Deep Neural Network to forecast the matching spoken language, these characteristics are first extracted.

Upon identifying a language, the web interface promptly displays the model's forecast beside it. This program can identify five of the most common Indian languages: English, Hindi, Bengali, Tamil, and Gujarati. Fast prediction and consistent classification performance for freshly uploaded voice samples are both ensured by integrating the learned model with the Django backend.

In addition to facilitating the management of audio files and prediction records, the created web app offers a structured framework for doing so. Django makes it easy to integrate the frontend, backend, machine learning model, and database via its modular design, which simplifies system maintenance and future expansions. Anyone with a web browser may now use the deployed framework to identify spoken languages automatically, without the need for any additional software or tedious feature extraction.

All things considered, the Django-based implementation turns the suggested research model into a workable real-world application that may be used for intelligent human-computer interaction, speech analytics, voice-enabled services, educational platforms, and multilingual communication systems. Moreover, the application's adaptable design for future inclusion of other languages, more sophisticated deep learning models, cloud deployment, and the ability to analyse speech in real-time without requiring substantial changes to the current setup.

VII. CONCLUSION

The five most common spoken languages in India—English, Hindi, Bengali, Tamil, and Gujarati—are all categorised here using a Deep Learning-based framework that makes use of Mel-Frequency Cepstral Coefficients (MFCCs). Data collection of spoken language, audio preprocessing, extraction of MFCC features, building of Deep Neural Network models, assessment of performance, and web-based deployment are all steps in the suggested methodology's methodical workflow. In order to correctly identify the language from input audio recordings, the suggested system learns language-specific speech patterns using representative acoustic characteristics and a multilayer Deep Neural Network architecture.

According to the results of the experiments, the suggested model can generalise well to new speech samples, as it obtains a classification accuracy of almost 90%. Based on the findings, it is clear that the model can accurately identify Gujarati, Hindi, and Tamil, and it makes good predictions for these languages. On the other hand, Bengali and English both share comparable acoustic features, thus there are only small classification mistakes between the two. In sum, a powerful method for identifying spoken languages across languages is available via the integration of MFCC feature extraction with Deep Neural Network classification.

Users may now submit audio recordings and get real-time language recognition results via an easy online interface thanks to the trained model's successful integration into a Django-based deployed web application, which enhances the practical usefulness of the proposed framework. Supporting intelligent multilingual communication systems, speech analytics, educational platforms, and

voice-enabled apps, the implementation demonstrates the practical usefulness of the proposed study.

Ultimately, the suggested architecture for spoken language identification provides an autonomous language recognition solution that is accurate, dependable, and scalable. Incorporating more Indian and foreign languages, bigger multilingual datasets, powerful deep learning architectures, cloud-based deployment for real-time speech processing applications, and future growth are all possible thanks to the system's modular design.

REFERENCES

1. J. Li, P. Zhan, and Y. Chen, “The HIT System for NIST 2015 Language Recognition Evaluation,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 5815–5819.
2. D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, “End-to-End Attention-Based Large Vocabulary Speech Recognition,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 4945–4949.
3. F. Richardson, D. Reynolds, and N. Dehak, “Deep Neural Network Approaches to Speaker and Language Recognition,” IEEE Signal Processing Letters, vol. 22, no. 10, pp. 1671–1675, 2015.
4. J. Gonzalez-Dominguez, I. Lopez-Moreno, H. Sak, J. Gonzalez-Rodriguez, and P. J. Moreno, “Automatic Language Identification Using Long Short-Term Memory Recurrent Neural Networks,” in Proc. Interspeech, 2014.
5. S. Safavi, M. Russell, and P. Jančovič, “Identification of Sparse Audio Tampering Using Distributed Source Coding and Compressive Sensing Techniques,” EURASIP Journal on Advances in Signal Processing, vol. 2013, no. 1, 2013.
6. P. Matejka et al., “Neural Network Bottleneck Features for Language Identification,” in Proc. IEEE Odyssey: The Speaker and Language Recognition Workshop, 2018, pp. 299–304.
7. L. Wan, Q. Wang, A. Papir, and I. L. Moreno, “Generalized End-to-End Loss for Speaker Verification,” in Proc. IEEE ICASSP, 2018, pp. 4879–4883.
8. M. Gelly et al., “Spoken Language Identification Using LSTM-Based Angular Proximity,” in Proc. Interspeech, 2018, pp. 272–276.
9. N. Dehak, P. A. Torres-Carrasquillo, D. A. Reynolds, and R. Dehak, “Language Recognition via i-Vectors and Dimensionality Reduction,” in Proc. Odyssey Speaker and Language Recognition Workshop, 2011.
10. F. S. Richardson, D. A. Reynolds, and N. Dehak, “A Unified Deep Neural Network for Speaker and Language Recognition,” arXiv preprint arXiv:1504.00923, 2015.
11. Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, “PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2880–2894, 2020.
12. A. Garain, P. K. Singh, and R. Sarkar, “FuzzyGCP: A Deep Learning Architecture for Automatic Spoken Language Identification from Speech Signals,” Expert Systems with Applications, vol. 168, p. 114416, 2021.
13. S. Revay and M. Teschke, “Multiclass Language Identification Using Deep Learning on Spectral Images of Audio Signals,” arXiv preprint arXiv:1905.03330, 2019.
14. K. S. Bhogale et al., “Effectiveness of Mining Audio and Text Pairs from Public Data for Improving ASR Systems for Low-Resource Languages,” arXiv preprint arXiv:2208.12666, 2022.
15. W. Dong, J. Xu, X. Zhang, S. Xu, and B. Xu, “Automatic Speech Recognition Without Transcription,” in Proc. International Conference on Computer Pattern Recognition, 2018.
16. G. Gelly and J. L. Gauvain, “Spoken Language Identification Using LSTM-Based Angular Proximity,” in Proc. Interspeech, 2017, pp. 2566–2570.
17. F. Xia, C. Gao, G. Liu, X. Zhu, Q. Liu, and L. Dai, “Phonetic Posteriorgrams Based Data Augmentation for CNN Acoustic Models,” IEEE Access, vol. 7, pp. 93168–93177, 2019.
18. G. Singh, S. Sharma, V. Kumar, M. Kaur, M. Baz, and M. Masud, “Spoken Language Identification Using Deep Learning,” Computational Intelligence and Neuroscience, vol. 2021, Article ID 5758123, 2021.
19. Y. Bengio, I. Goodfellow, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
20. S. O. Haykin, Neural Networks and Learning Machines, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.
21. P. Ramesh Babu, “Automated Student Document Analysis and Query System Using OCR and Conversational AI,” pp. 358–366, 2026.
22. P. Ramesh Babu, “Reliable and Secured Banking Management System Through Machine Learning,” Journal of Applied Science and Computations, vol. 6, no. 4, pp. 377–382, Apr. 2019.