AUTOMATING LEGAL TEXTS: DEEPLEARNING FOR SPEECH GENERATION AND TRANSCRIPTION

Nuka Nandan Vignesh – 21BRS1473 , Narra Sathwik Redyy - 21BRS1548
Dr.Sivaranjani N
School of Computer Science and Engineering
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI

V-NEST
VIT CHENNAI STARTUP AND RESEARCH FOUNDATION
WE NURTURE YOU TO FLY HIGH

MOTIVATION / INTRODUCTION

- This system transforms legal judgment transcription into automated work so it requires less effort and time for handling complex legal documents.
- The system employs BERT to extract important judgment data which enables it to separate core sections containing facts and legal arguments and judge decisions for precise transcription.
- A customized legal data set from IndianKanoon feeds into the system to create a transcription model specialized for Indian legal judgment expression.
- The system converts processed legal content text through Text-to-Speech (TTS) models because these produce high-quality speech which enables audio accessibility.

OBJECTIVES

- Automate legal judgment transcription using deep learning.
- Use BERT and rule-based systems to extract content (400–800 words).
- Convert extracted text to speech using gTTS. Transcribe audio using Whisper, Vosk, and Wav2Vec 2.0.
- Evaluate results using WER, CER, and BLEU Score.

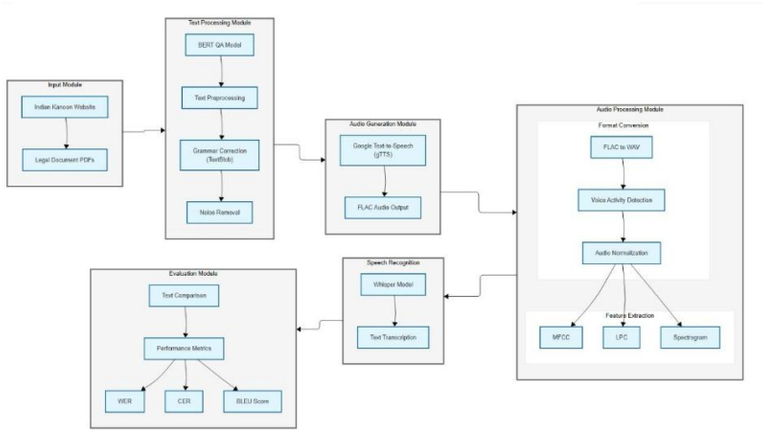
SCOPE OF THE PROJECT

The system enhances legal document accessibility by improving transcription speed and accuracy. It supports multilingual transcription and performs well in noisy environments. Real-time transcription capabilities benefit legal professionals. Additionally, it addresses scalability, structure, and reduces errors in manual processes.

METHODOLOGY

- Legal judgments from IndianKanoon.
- Using BERT + Rule-based fallback (min. 400 words).
- Using gTTS to convert judgment to .flac audio.
- VAD + Normalization + Feature Extraction (MFCC, LPC).
- Speech to Text using Whisper, Vosk, and Wav2Vec 2.0.
- Metrics: WER, CER, BLEU Score.

ARCHITECTURE



Input Module: This module takes legal documents from Indian Kanoon in PDF format. These files typically include judgments, case details, and procedural content, forming the foundation for further processing.

Text Processing Module: A BERT-based Question Answering model extracts judgment-related content. If extraction is below 400 words, rule-based and fallback extraction methods are applied. The text is then cleaned, grammatically corrected, and denoised to prepare it for audio generation.

Audio Generation Module: The processed legal text is converted into speech using Google Text-to-Speech (gTTS). The output audio is stored in FLAC format to maintain high quality, which is crucial for accurate transcription.

Audio Processing Module: This module prepares the audio for transcription. It converts FLAC to WAV, applies Voice Activity Detection (VAD) to remove silence, normalizes volume, and extracts key audio features like MFCC, LPC, and spectrograms.

Speech Recognition Module: The Whisper model by OpenAI is used to transcribe the processed audio back into text. It handles various speech complexities such as legal terminology, accents, and ambient noise effectively.

Evaluation Module: This module compares the original extracted text with the transcribed output. It uses standard metrics like Word Error Rate (WER), Character Error Rate (CER), and BLEU Score to assess the performance of the transcription system.

RESULTS

Model	WER	CER	Bleu Score
Wav2Vav 2.0	63.06	33.46	0.28
Whisper	16.07	5.99	0.76
Vosk	43.9	29.4	0.52

CONCLUSION

This research project tested three speech-to-text models namely Whisper, Wav2Vec 2.0 and Vosk to evaluate their effects on legal text extraction performance. The performance of Whisper surpassed all other models through its minimal Word Error Rate and Character Error Rate while producing the highest BLEU score that demonstrated its superior translation quality and accuracy.

CONTACT DETAILS

MAIL ID – nukanandu1008@gmail.com
MOBILE NO - 7842908036

REFERENCES

1. Lyra A., Barbosa C.E., Santos H.S., Argôlo M., Lima Y., Motta R., Souza J.M. Automatic Transcription Systems: A Game Changer for Court Hearings. Federal University of Rio de Janeiro, 2024.
2. Prasad R., Nguyen L., Schwartz R., Makhoul J. Automatic Transcription of Courtroom Speech. BBN Technologies, 2025.
3. Brown J., Tanaka H., Patel D. Domain-Adapted Speech Models for Legal Transcription. Speech Communication Journal, 2024.
4. Zhang F., Liu X., Kapoor R. Adversarial Robustness in Legal Speech Recognition Models. Neural Networks Journal, 2024.
5. Hassan M., Oliver T., Chen B. Enhancing STT Accuracy with Legal Ontologies. Expert Systems with Applications, 2024.
6. Adams R., Nguyen P., White J. Ethical Challenges in AI-Based Legal Transcription. AI & Ethics Journal, 2023.
7. Wang Y., Gao J., Chen J. Deep Learning Algorithm for Judicial Judgment Prediction Based on BERT. IEEE, 2020.