Group 4

Machine Learning

Noise Reduction

Final Project
Digital Signal Processing & Machine Learning

Gabriel Ruben Weslie
2802404204

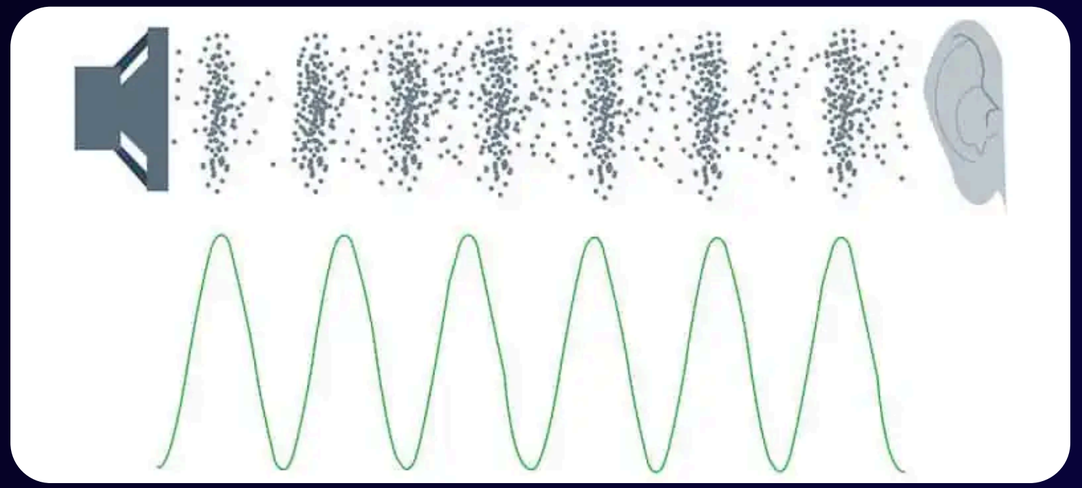
Takeshi Gobstan Lee
2802540471

Rizky Kresnanto Dananjaya
2802479142

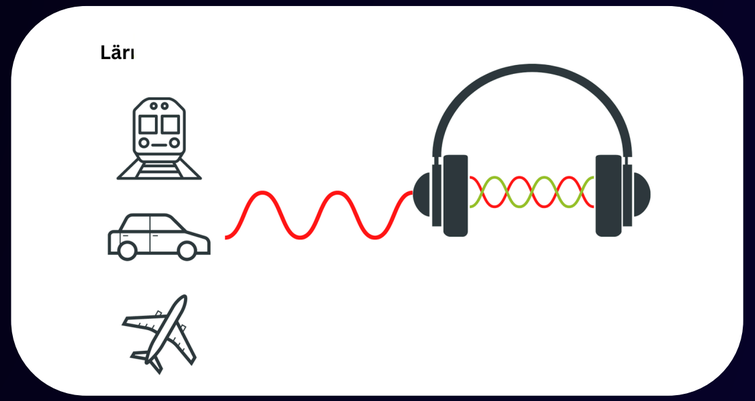
Jordan Elishua Wibowo
2802466865

Background

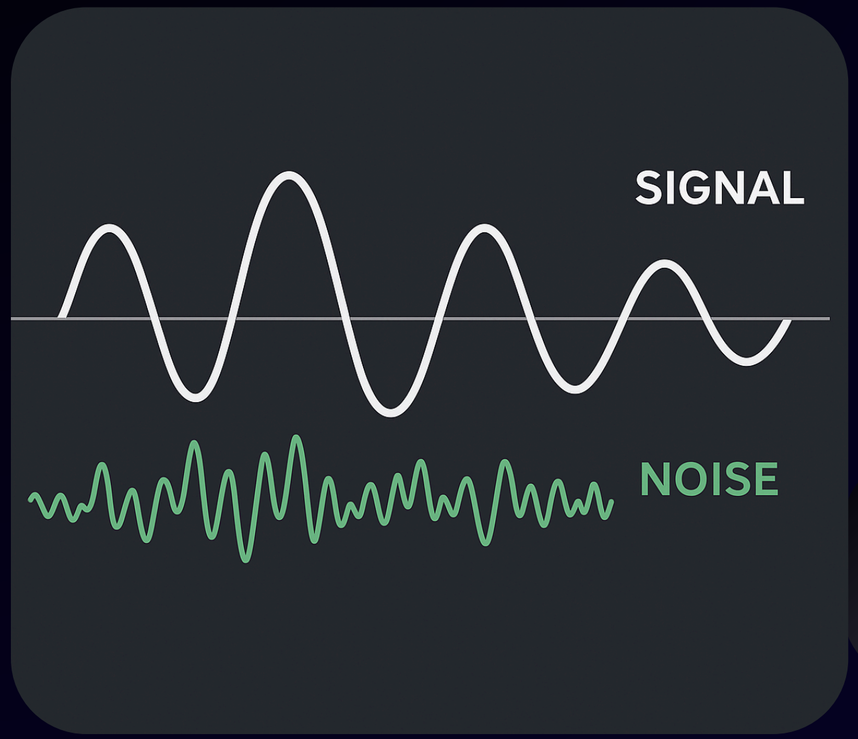
To solve noise disruption that often reduces audio clarity, this project combines the Short-Time Fourier Transform (STFT) with a Long Short-Term Memory (LSTM) model. Unlike classical methods, LSTM is able to capture temporal correlations in audio signals, allowing it to recognize human speech patterns and separate them from dynamically changing noise more accurately.



Objectives



1. To evaluate the system's adaptability to various types of noise.



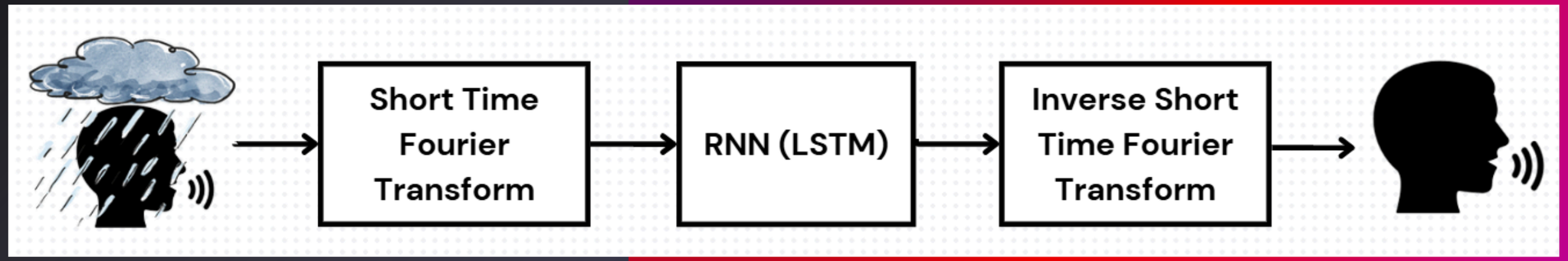
2. To improve signal quality in a measurable way.

3. To implement a Long Short-Term Memory (LSTM) model to automatically predict clean signals from noisy inputs.

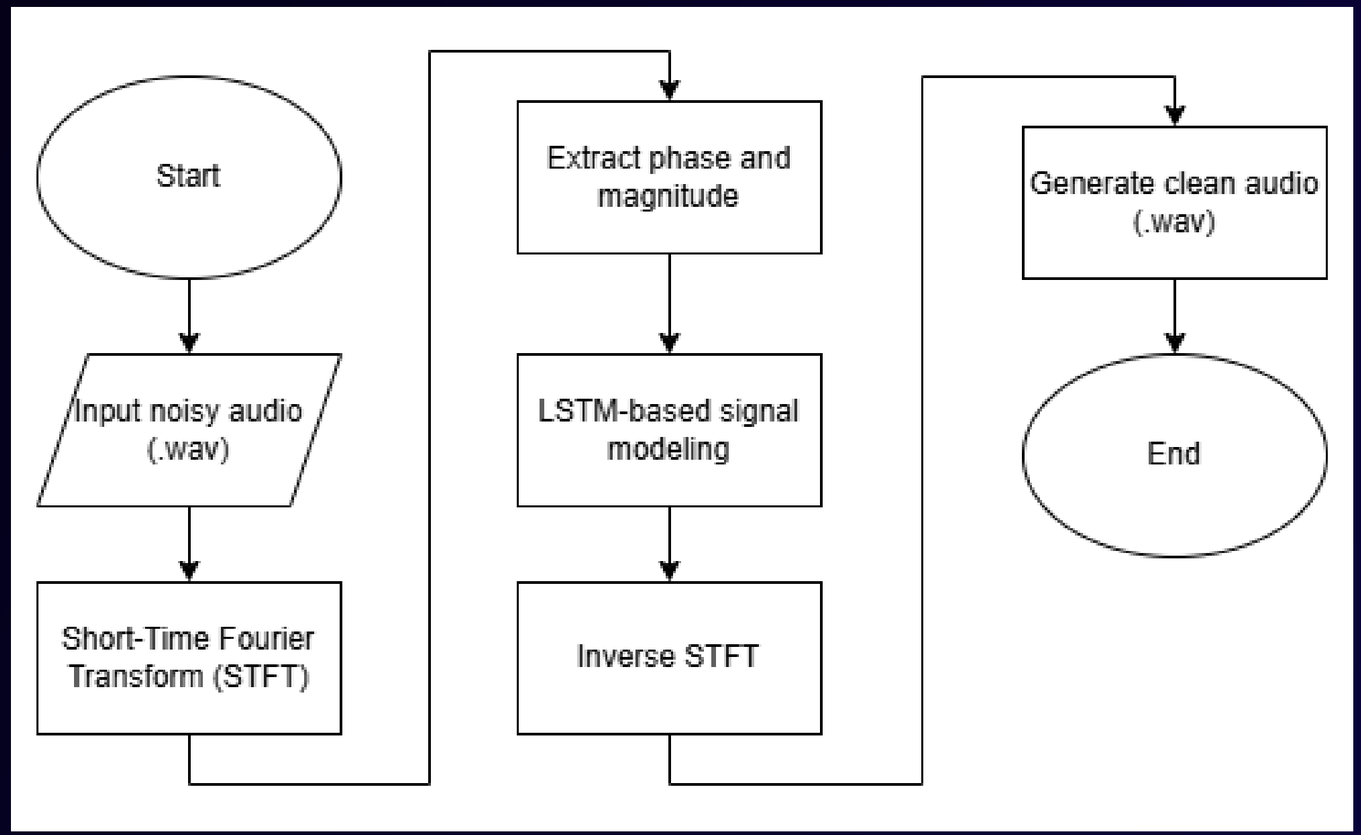
Dataset

Valentini-Botinhao, Cassia. (2017). Noisy speech database for training speech enhancement algorithms and TTS models, 2016 [sound]. University of Edinburgh. School of Informatics. Centre for Speech Technology Research (CSTR). https://doi.org/10.7488/ds/2117.

Methodology



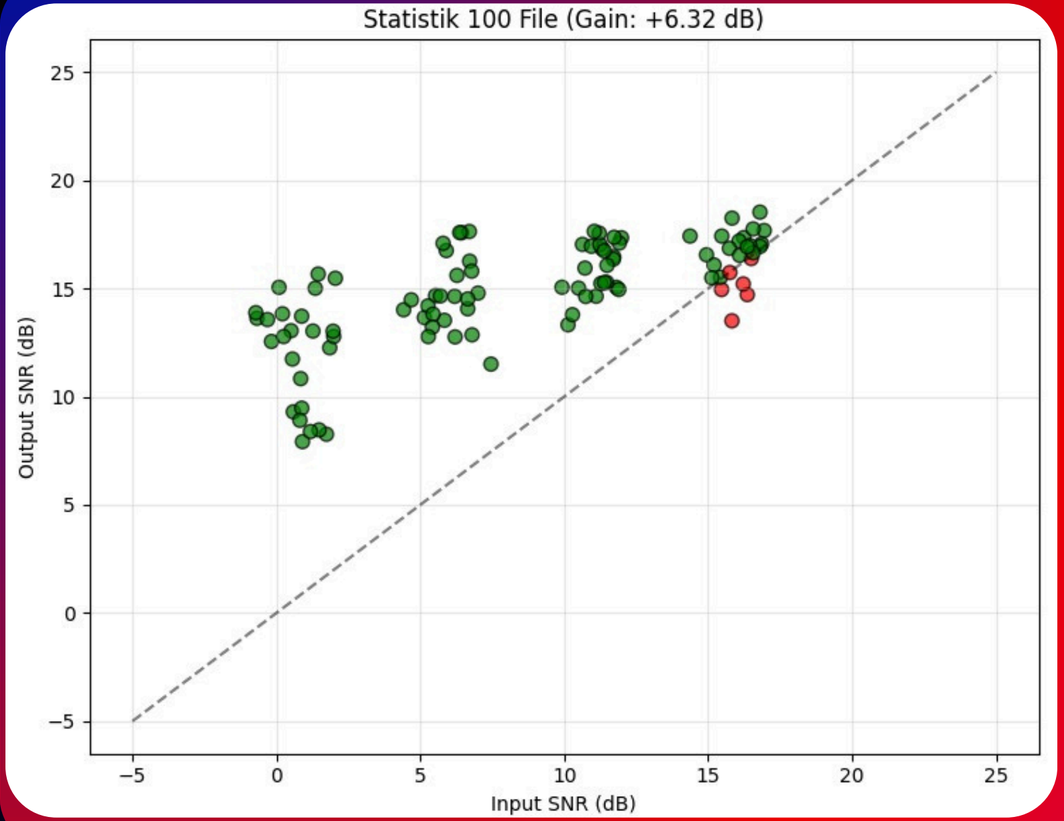
Flowchart

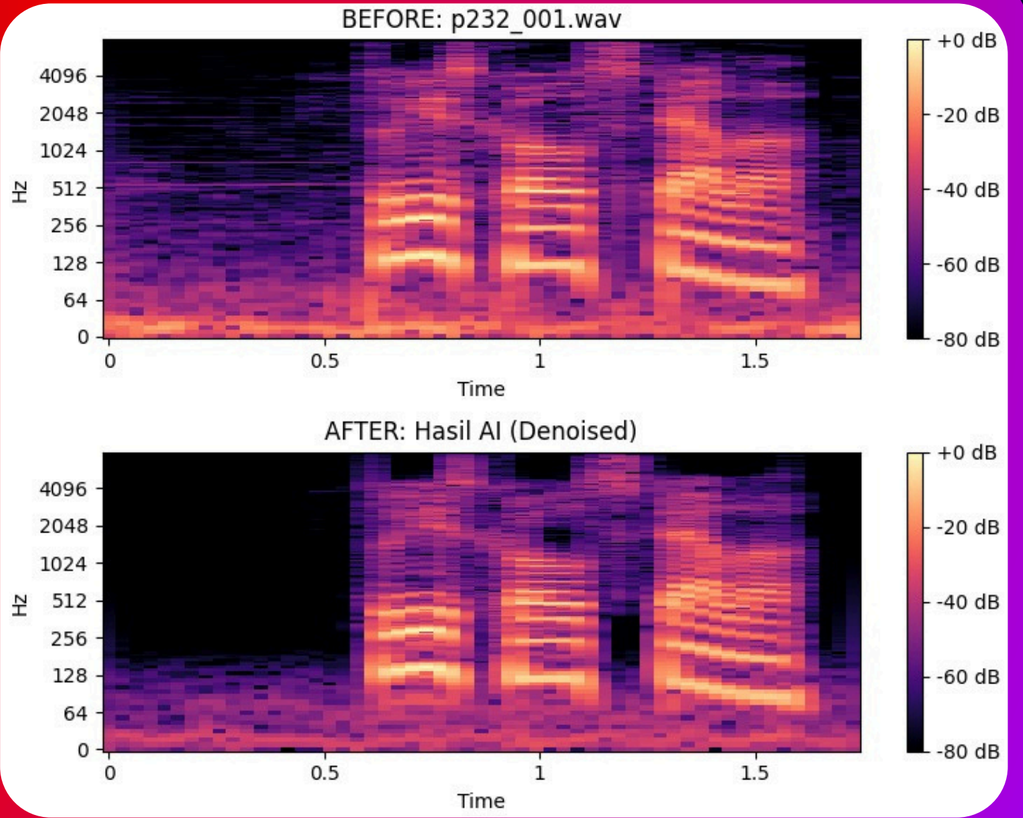


Result :

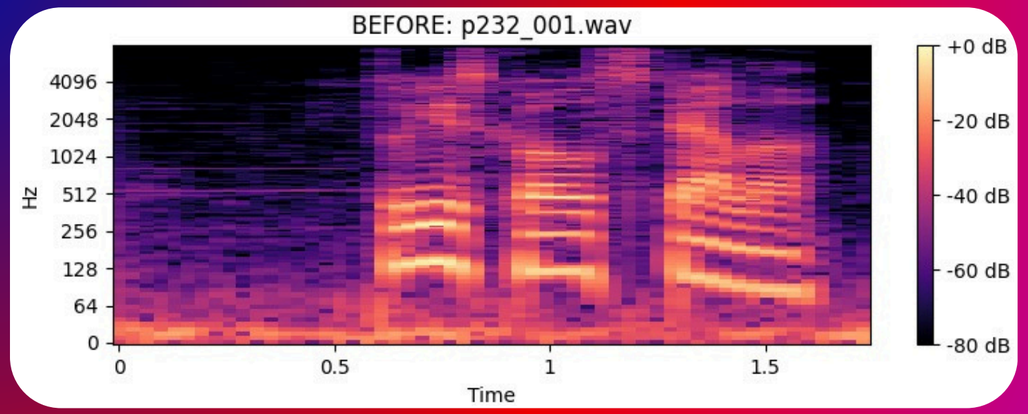
Human speech without rain noise

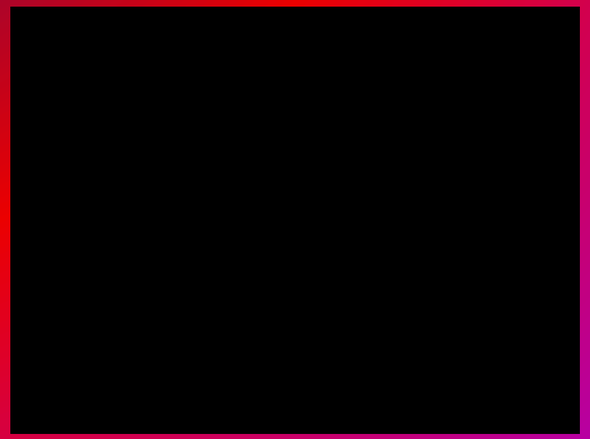
Result

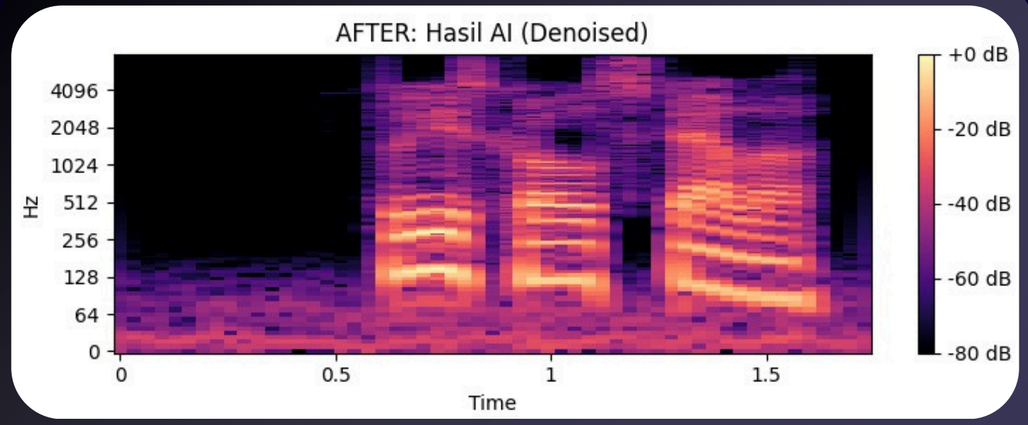

Statistik 100 File (Gain: +6.32 dB)
Output SNR (dB)
Input SNR (dB)
25
20
15
10
5
0
−5
−5
0
5
10
15
20
25


BEFORE: p232_001.wav
Hz
4096
2048
1024
512
256
128
64
0
0
0.5
1
1.5
Time
+0 dB
-20 dB
-40 dB
-60 dB
-80 dB
AFTER: Hasil AI (Denoised)
Hz
4096
2048
1024
512
256
128
64
0
0
0.5
1
1.5
Time
+0 dB
-20 dB
-40 dB
-60 dB
-80 dB

Result


BEFORE: p232_001.wav
4096
2048
1024
512
256
128
64
0
Hz
0
0.5
1
1.5
Time
+0 dB
-20 dB
-40 dB
-60 dB
-80 dB




AFTER: Hasil AI (Denoised)
4096
2048
1024
512
256
128
64
0
Hz
0
0.5
1
1.5
Time
+0 dB
-20 dB
-40 dB
-60 dB
-80 dB

Conclusion

By using an LSTM-based temporal modeling approach, audio filtering becomes more effective. The proposed system improves audio quality, achieving an average SNR increase of +6.32 dB. The model is able to separate human speech from background noise (such as rain or machinery) by analyzing temporal context between-frames.

Group 4

Thank You

Machine Learning

-