**Spectogram Analysis Document**

**Context:**

* The visual depiction of a signal's intensity over time at various frequencies included in a particular waveform is called a spectrogram.
* It is depicted as a two-dimensional graph in which time is shown along the horizontal axis, frequency along the vertical axis, and the intensity or color of each point on the graph denotes the amplitude of the frequency components at that specific moment.
* Dark blue hues signify low amplitudes, whereas brighter hues up to red indicate greater (or louder) amplitudes.
* By applying the Fast Fourier transform (FFT) to the speech signal, which creates a time-frequency representation, it is calculated from the spoken signal.
* The signal is broken up into smaller pieces in order to find the frequencies at a certain point, and FFT is then applied to the speech waveform for each chunk.

**Observations:**

* We employed a TESS Toronto emotional speech set data that had 6 emotions namely Angry, Disgust, Neutral, Sad, Pleasantly Surprised & Happy with 400 audio samples each
* Each waveform-spectrogram combination created for the input voice signals was analyzed and used for prediction using the LSTM model
* Our model generated probabilistic predictions for the six various emotions, which are then employed to get overall prediction scores for these emotions
* The model's predictions for each unique spectrogram serve as support for updating the belief values for all emotions, such as in Anger audio samples, the Amplitude and the Frequency of words was too high whereas in Sad audio samples both the aforementioned characteristics was low.
* In the present situation with six classes, there is a fair possibility that the specific emotion may be predicted successfully based on the gathered data from various spectrograms if around 25–30% of predictions for the audio files are done correctly
* By calculating mean predictions from the obtained data, the predictions are pooled to establish the probability for certain emotions and we see that most of the test data contains either neutral, happy or confident emotions

**Scope for Improvement:**

* The suggested framework has to be improved in order to identify all emotions robustly and effectively
* To boost SER performance even further, we want to employ additional data and very complicated models