This study utilizes EEG signals for an objective diagnostic approach using deep learning. We applied Continuous Wavelet Transform (CWT) for classification. Then Vision Transformer (ViT) distinguished between control and alcoholic subjects using 64-channel UCI EEG data sampled at 256 Hz. the ViT model worked on wavelet-based heatmaps. the Vision Transformer demonstrating particularly strong performance, achieving an accuracy of 94.87%, precision of 94.76%, and recall of 95.1%.

1) *Wavelet Transform:*

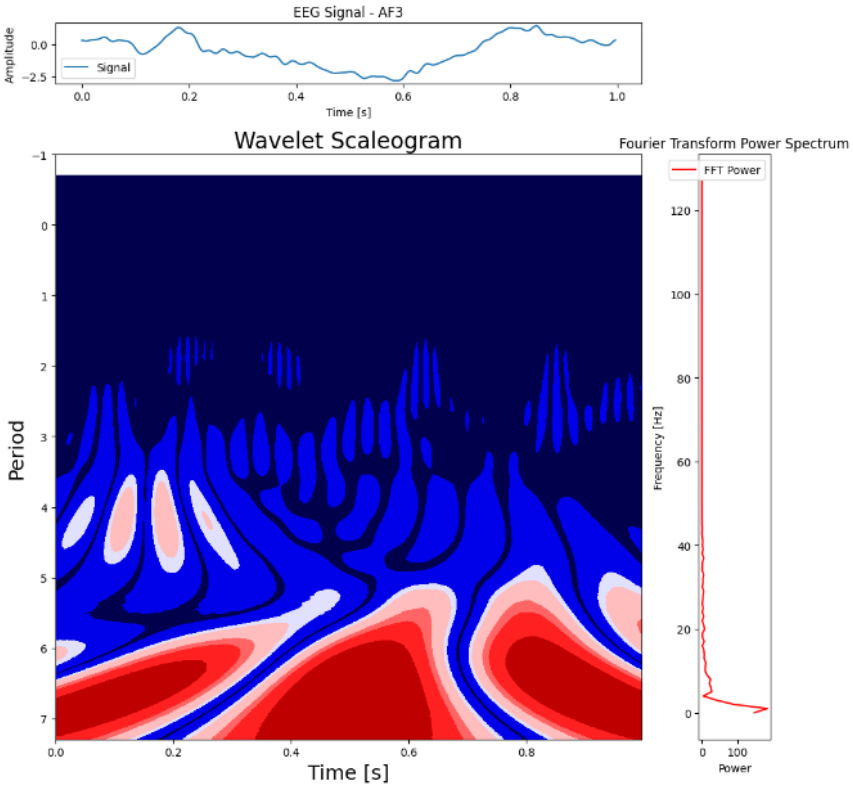
Signal analysis based on a family of concentrated wavelets in the time and frequency domains captures the essence of certain scenarios, such as zero means and finite energy, and their behaviour [16]. Mathematically the total energy of wavelet function is expressed as:

$$\int_{-\infty}^{\infty} |\varphi(t)|^2 \, dt < \infty \qquad (2)$$

In our experiment we used the CWT for each 61 channels of the EEG signals for both alcoholic and non-alcoholic. The CWT of a signal is represented as:

$$CWT_x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\varphi^*\left(\frac{t-b}{a}\right) dt \qquad (3)$$

The CWT output here is the 2D heatmaps, known as Scalo grams, where the x-axis and y-axis represent time and scale, respectively. The intensity of the heatmap represents the magnitude of wavelet coefficients which reflect the signal wavelet energy at different scales and transition. Generated heatmap after CWT for a single channel is shown in figure 4.



In this case, the x-axis covers 0 to 1 second, while the y-axis represents the period (inverse of frequency) for the AF3 channel sensor values. The period in the wavelet transform links each scale to time intervals of signal oscillations. The Fast Fourier Transform (FFT), shown alongside the wavelet heatmap, highlights that the wavelet provides temporal information, which is absent in FFT analysis. The CWT parameters which we implemented are given below in table II.

TABLE I. PARAMETERS USED FOR CWT

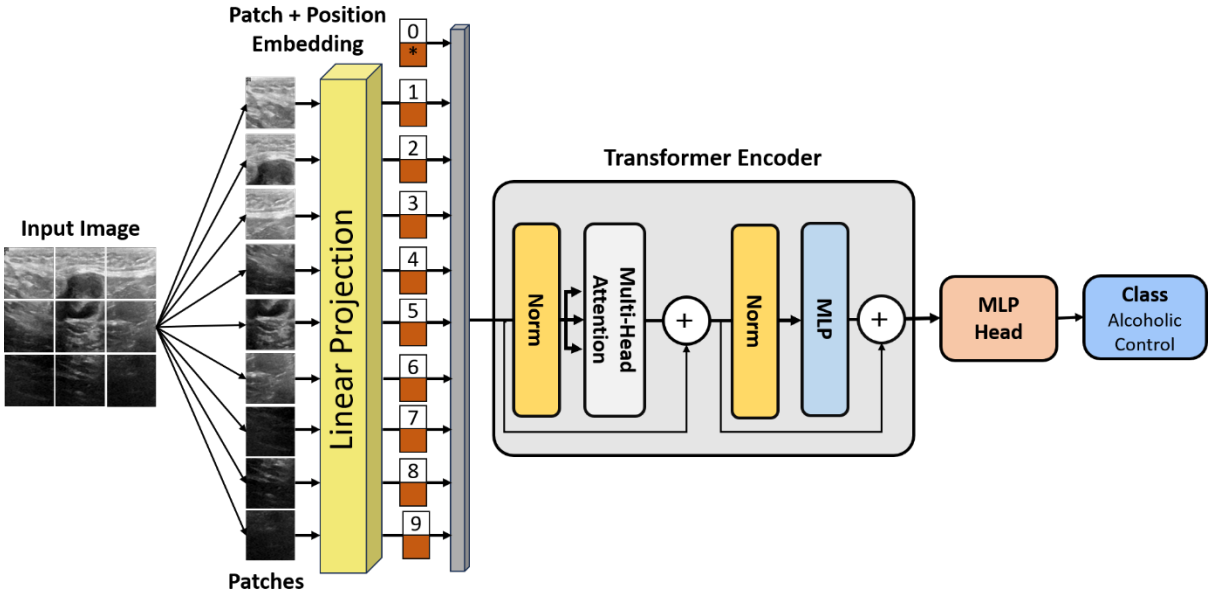
Parameters	***Conditions***
Wavelet	'Morlet'
Scale	Lower = 1, Higher = 255
Sampling interval	0.5
Contour Level	base-2 logarithm

We organized our dataset such that the signals are sampled at 256Hz for a time period of 1 sec. The time slot was divided into a total row of 256 ranging from 0 to 255 where the column indicates the sensor values of all the 61 electrodes minimizing the noise for an individual person.

In the context of CWT feature extraction, the train-test file was merged, and 488 images (8 subjects & 61 channels) for each alcoholic and control group was expected. But for some duplicate values those were aggregated by mean values which resulted in 468 images for alcoholic and 456 for control group.

2) *Vision Transformer for Wavelet Transformation*

The Vision Transformer (ViT) architecture divided the input image into patches, treated each embedded patch as a word in natural language processing, and used self-attention modules to figure out how these patches relate to one another. Compared to CNNs, ViTs perform better on image classification tasks because they include more global information and stronger skip connections.



The proposed workflow is picturized below in figure 7 for better understanding. Binary classification of alcohol and control group has been made to feed to the ViT model with the image size being (200X200) and the total number of images we got was 924. After taking 8X8 patches of the input image, it was fed into the embedded patch layer where it had the position embeddings. Then it passed through the normalization layer and travelled across a multi-head attention layer and this process continued two times. Note that there was skip connection to add the two normalization layers. MLP layers consist of 'ReLU' layers and 'Adam' optimizers. Afterwards, the output of this layer, the logits,

was used to calculate the loss function as well as the final probability function, which classified the images into their respective classes.

The parameters used for the transformer model are shown below in table IV.

TABLE II. PARAMETERS USED IN VIT MODEL

Parameters	***Conditions***
Patch Size	8 X 8
Activation function	'relu'
Optimizer	'Adam'
Learning Rate	0.001
Max. iterations	14
Epochs	16
Batch size	64