Novel Welch-Transform based Enhanced Spectro-Temporal Analysis for Cognitive Microsleep Detection using a Single Electrode EEG

Jash Shah, Amit Chougule and Vinay Chamola, *Senior Member, IEEE* and Amir Hussain, *Senior Member, IEEE*

***Abstract*—The growing demand for semi-autonomous human-machine systems has led to an increased requirement for human fatigue detection. Direct and invasive approaches for microsleep detection include cognitive computing methods using Brain-Computer Interfaces (BCI). The contextual integration of multi-channel or heterogeneous signals for sleep staging remains a formidable challenge. In addition, the cost of acquiring many signals is significantly higher than that of acquiring a single signal. Consequently, researchers have recently attempted to utilize single-channel EEG over multi-channel acquisition systems for sleep staging. The Fast Fourier Transform (FFT) has been widely used in previous research findings for spectral analysis of complex time-series data streams. In contrast to the FFT, we utilize here, for the first time, the Welch Transform which can give higher stability in noise reduction for spectral analysis. Specifically, we provide a novel method to implement the short-time Welch transform (STWT) as an enhanced technique for the spectro-temporal analysis of single-electrode EEG signals. Further, our proposed model utilizes attention-based spatial and channel-wise inter-dependencies using a one-dimensional causal convolutional neural network (CNN) to extract contextual features automatically. Finally, we demonstrate an end-to-end proof of concept for our data extraction, adaptive data resampling, manual feature extraction, and deep-neural network-based modeling architecture. Comparative simulation results using the benchmark, maintenance of wake-fullness test (MWT) dataset for microsleep detection during automobile transportation, show that our proposed end-to-end system, utilizing novel STWT-based enhanced spectro-temporal analysis, outperforms current state-of-the-art methods, delivering 95% and 89% test accuracy for the case of temporal and spectral data inputs, respectively.**

***Index Terms*—Adaptive Resampling, Cognitive Microsleep Detection, Convolutional Encoder-Decoder, Multi-Scale Feature Extraction, Short Time Fourier Transform, and Welch's Transform.**

I. INTRODUCTION

Autonomous Vehicles (AV) have evolved to support the enlistment of a new lifestyle by assisting a variety of actions that were not possible earlier in manually controlled vehicles. This changed lifestyle has shifted the need for a driver from actually "driving" the car to just "supervising" it. At present, all self-driving cars are partially autonomous. All the self-driving vehicles are in level 2 of driving automation. Level two driving automation is an advanced driver assistance system (ADAS) that allows vehicles to regulate steering and acceleration. However, there are instances when the driver must take control of the vehicle. Recent AV research has focused on studying drivers' cognitive demands while engaged in secondary activities. According to various research articles, sleeping, observing the scenery, listening to music, and conversing on the phone are drivers' most common secondary tasks [1]. However, in Level 2 Autonomy, the driver's inattention or drowsiness can completely redirect the driver's attention from the road causing menacing crashes and accidents or, in the worst-case scenario, deaths [2]. AAA Foundation performed analysis for traffic safety; out of all various types of on-road accidents, 16.5% of accidents in the US are due to microsleep [3]. Another National Highway Traffic Safety Administration report demonstrates that 2.5% of fatal crashes and 2% of injury crashes, which is 6,000 fatal crashes yearly, occur due to drowsy driving [4]. Truck drivers who unexpectedly fall asleep while driving cause 30 to 50 percent of accidents in Germany [5]. Due to the immense demands on experimenters and subjects, current AV research has yet to examine the falling asleep situation. American Automobile Association for Safe Driving and a traffic investigation safety group have released several studies [6]. Microsleep is one of the primary reasons that causes driver inattention, causing around 16.5% of fatal car

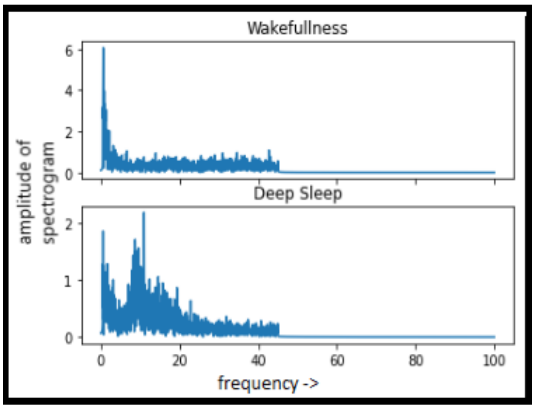

Fig. 1: Spectograms of Alpha Wave (Wakefulness) and Theta Wave (Deep Sleep) of the same subject. The frequency in Hz is shown by the X-axis, while the Y-axis shows the spectral power.

Jash Shah and Amit Chougule are with the Department of Electrical & Electronics Engineering, BITS-Pilani, Pilani Campus, 333031, India. (e-mail: jashshah0801@gmail.com, amitchougule121@gmail.com).

V. Chamola is with the Department of Electrical and Electronics Engineering & APPCAIR, BITS-Pilani, Pilani Campus, 333031, India. (e-mail: vinay.chamola@pilani.bits-pilani.ac.in).

Amir Hussain is with School of Computing, Edinburgh Napier University, Scotland, UK (email: A.Hussain@napier.ac.uk)

crashes—monotonous tasks, including driving and watching. In a 2012 study by G.Poudel and others [7], participants were invited to play a 50-minute computer game in which they used a joystick to follow a dot around the screen. Researchers tracked eye movement and brain activity for indicators of tiredness during that time. The study found that participants had an average of 79 microsleep events, with some extending up to six seconds.

A microsleep is a brief sleep episode lasting between one and thirty seconds. It occurs when there are abrupt transitions between the state of "awakeness" and "sleepiness". In terms of Electroencephalography frequency, microsleep relates to the frequency shift in EEG signal where 4-7 Hz theta wave (indicative of microsleep) activity substitute the 8-13 Hz alpha wave (indicative of wakefullness) pattern [8, 9], as shown in Fig 1. This shift of the brain waves is an indicative of the onset of sleep in an individual. People who have these experiences may fall asleep without even recognizing it. Drowsiness, difficulty focusing, heavy eyelids, blank staring, and yawning are some signs of microsleep. This odd state of awareness is distinguished by brief bursts of sleep that occur while a person is awake, frequently while their eyes are open and they are either sitting upright or executing a task. Microsleep is characterized by brain sections going offline for a few seconds while the rest of the brain remains awake [10]. From the driver's perspective, a micro-sleep can negatively impact response time and outcomes comparable to or even worse than driving with dangerous drunk and driving cases. Such induced inattention while driving can be dangerous to surround vehicles and pedestrians and cause significant accidents claiming people's lives. Those who are really drowsy during the day frequently experience microsleeps [11, 12]. Microsleep can cause attention gaps that make it harder to recognise and respond to important cues and activities. In a multi-parametric attention tracking system and poor driving-related performance, microsleeps have also been connected. Therefore, it is crucial to develop ways to detect and avoid it. Sleep deprivation, narcolepsy, and sleep apnea can cause microsleep. Still, it typically occurs when we lack adequate sleep or perform repetitive tasks. When a person is driving, handling construction equipment, or performing other safety-critical tasks that are repetitive in nature, microsleep becomes a serious problem. A study on Sleep Disorders was provided by the National Commission [13], which indicates the cost of Sleep-Related Accidents and the role of drowsiness in adding to the total number of accidents. They analyzed and identified the number of car accidents caused by sleepiness in their research. The overall cost of sleep-related accidents was reported to be between $ 43.15 billion and $ 56.02 billion.

This paper suggests a non-invasive and active microsleep detection model using a single-electrode EEG signal. The proposed model aims to detect microsleep episodes by analyzing the EEG signal collected from a single electrode, without the need for invasive methods.The suggested BCI-induced brain plasticity replicates human cognitive neural responses by translating brain activity signals into a computerized command. Our contributions in this research work are as follows:

- We present a method for automated analysis of single-channel EEG-based microsleep detection using a Multi-Layer Neural Network architecture.
- The proposed model utilizes Multi-filter convolutions precisely captures and learns temporal as well as spectral attributes of the signals in real time.
- In the proposed approach, the adaptation of Spectro-Temporal Analysis using Short Time Welch's Transform shows improved outcomes.
- Finally, the paper also emphasizes an improved Sampling Strategy to address the imbalances in the training dataset.

This paper is subdivided into the five sections listed below: The I section represents the idea of microsleep, as well as key concepts and the significance of this study. The essential literature regarding previous comparable studies is included in Section II. Section III talks about elementary data pre-processing. Section IV represents the preliminary background. Section V contains our proposed framework in depth. Section VI presents the experimental results based on our proposed framework. Section VII summarizes our results and further analyses their demographics.

Dataset Name	**Electrode**	**Study Type**	**Sample Rate**
SleepEdfX [14]	Pz-Oz	Monopolar Reference Study	100Hz
Drowsiness-DB [15]	Pz-Oz	Monopolar Reference Study	100Hz
MWT [16]	O1-M2 and O2-M1	Measure of posterior activity	200Hz
HMC [17]	O1-M2 and O2-M1	Measure of posterior activity	256Hz
CCNL [18]	O1-M2 and O2-M1	Measure of posterior activity	200Hz
CCNL [18]	C4-M1	Measure of central activity	200Hz
HMC [17]	C4-M1	Measure of central activity	256Hz
CCNL [18]	F4-M1	Measure of frontal activity	200Hz
HMC [17]	F4-M1	Measure of frontal activity	256Hz

TABLE I: Details of datasets

II. Literature Survey

Our work of mimicking the brain functions and processing them through a computing unit has been inspired by the study in cognitive AI. With cognitive AI, machines get trained to be able to think, reason, and make decisions much like humans do![19–21] Although existing research work had restricted scopes due unavailability of real-world datasets and problematic experimental conditions, researchers yielded results using indirect methods like observing subject's typical sleep patterns in the vehicle [22, 23, 23], or experimented with decreased cognitive response time using alcohol [24, 25] or conducted repetitive tasks to elicit tiredness [26, 27, 27]. When it comes

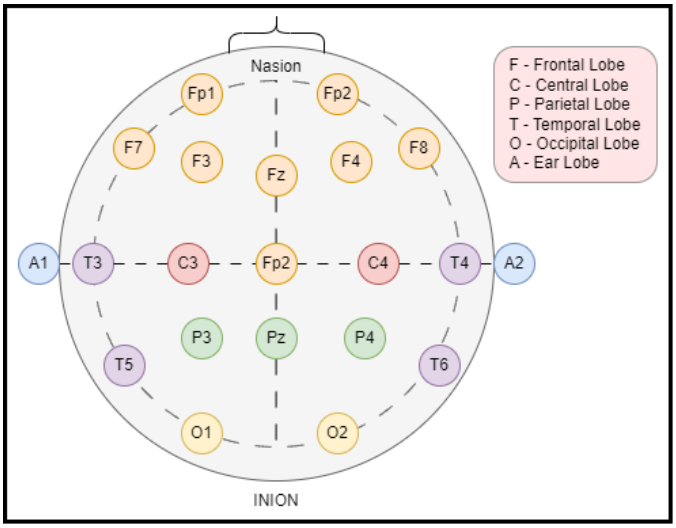

Fig. 2: EEG electrode positions according to the 10-20 System.

to sleep, there is no such thing as a typical state of being. Instead, sleep is divided into several stages, each of which may be differentiated by the pattern of brain wave activity that occurs throughout each stage [28, 29]. EEG may be used to examine these variations in brain wave events, distinguished by the amplitude and frequency of the brain waves [28, 30, 31]. The authors had reviewed the datasets from SleepEdfX [14], Maintenance of Wakefulness Test (MWT) [16], Drowsiness-DB [15], Haaglanden Medisch Centrum Data (HMC) [17] and Computational Clinical Neurophysiology Laboratory (CCNL) [18], the details of which have been summarized in Table I. To uniformly review these datasets that were retrieved at different sample frequencies, we upsampled the data using poly-phase filter. This involves aliasing of data. After that, a low pass FIR filter is applied and finally the data is down sampled back to the required frequency. This method does not need the data to be periodic, which counters the effect non-uniformity issued due to mastoid electrode.

Detailed research conducted by Michael S. Aldrich [32] addressed 424 subjects suffering from sleep problems such as sleep apnea, narcolepsy, other diseases with excessive sleepiness, and sleep disorders. This experiment indicates that people affected by hypersomnia condition (Excessive sleepiness) have a 3-7 % chance and Apneic & narcoleptic patients are responsible for 71 % of all sleep-related accidents per year. Although narcoleptics had the highest rate of sleep-related accidents, apneic were engaged in more sleep-related accidents because of their higher frequency. In yet another research work on the effect of sleepiness, the Italian highway vehicle accidents (1993 to 1997) by Garbarino [33] used a 24-hour sleep curve to highlight the relationship between accidents (sleep-related or not) and drowsiness. This article also used linear regression techniques to study the relationship between sleep-affected and non-sleep-affected accidents.
Current non-EEG-based microsleep monitoring methods employ facial recognition-based techniques involving keypoint detection and feature extraction [34, 35] and algorithms to track and monitor eye movement [36, 37]. However, these approaches could be more robust, and their efficiency varies greatly per surrounding conditions such as low light and foggy environments. EEG signal offers various applications[38], from stress detection [39], to various driving skills assessments [40]. Integration of CNN and RNN-based Deep Learning Approaches and their variants have been widely used in this domain. Kweon *et al.* [41] proposed U-Net [42] based automated microsleep identification system that relies on a vehicle driving simulation model and Night-sleep EEG. Aldrich *et al.* [43] suggested a method for interpreting EEG sleep stage scores that employs a deep convolutional neural network. The author uses multitaper spectral analysis to create visually recognizable representations of sleep patterns by feeding EEG data into a network trained to perform visual identification activities. In this method, Multiple convolutions were used to begin the preprocessing, followed by global average pooling and then Spcetrography. Phan Huy, *et al.* [44] suggested a state-of-the-art approach that uses Hierarchical Recurrent Neural Network for Automatic Sleep Staging. The network has loaded a sequence of many epochs and classifies various sequences into labels. The network is made up of a filter-bank unit meant to perform frequency-domain filtration for processing and an attention-based recurrent layer created for short-term sequencing modeling at the epoch stage. Mousavi *et al.* [45] presented a Deep Learning-based approach for Automatic Sleep Stage Ranking. A single-channel EEG signal is employed in this approach. Convolutional neural networks (CNNs) are also utilized to obtain time-invariant properties, frequency information, and a sequence-to-sequence model to acquire the complex and long short-term context correlations among sleep epochs and scores. Complicated models like [46] involves multiple residual blocks followed by an Adaptive Channel Fusion module (similar to U-Net - simple concatenation in the z-axis). Finally, a Hidden Markov Model is utilized to analyze the probability of occurrence of the current state, given the previous states and the effect of current states due to the future states and probabilities.

The idea of transfer learning is familiar in the domain of deep learning. Many articles include the improvement of previously trained models using transfer learning followed

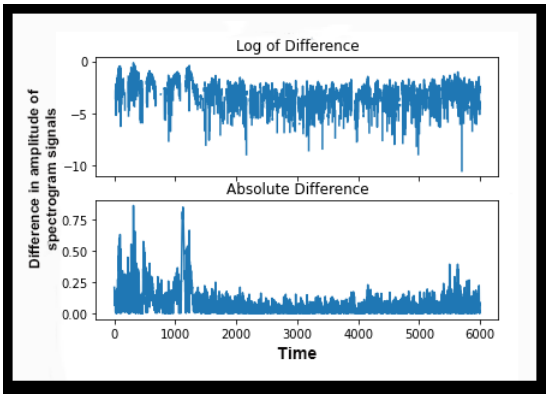

Fig. 3: Magnitude of difference between the signal, before and after passing the notch filter. (Here, X-axis denotes the time, and Y-axis is Difference in amplitude of spectrogram signals)

by hyper-parameter tuning and preceded by improved feature selection methodologies. Vilamala Albert [47] used the famous VGG neural architecture to model the last three layers being fine-tuned. Initially, the pre-processing involved multiple convolutions, followed by global average pooling, and finally followed by Spcetrography. This is further classified using ANNs. Sensitivity Analysis involved summed averages of differentiation of loss with respect to that particular input feature. Adversarial training models also showed good performance along with Discriminative Models. In the article, [48], the authors employed adversarial training and spectral regularisation to provide sleep-staging DNNs with strong noise resilience. The Lipschitz constant theory is used to eliminate network noise.

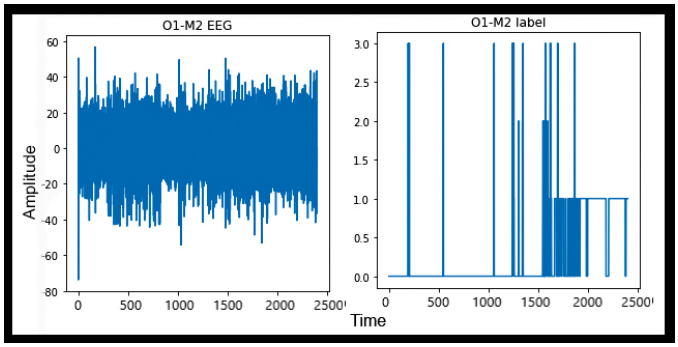

Fig. 4: Visualization of sample data (The x-axis denotes time in seconds. Y-axis showcases the signal's amplitude)

III. Elementary Data Pre-processing

EEG electrodes can be placed at multiple positions, out of which, the 10-12 electrode positioning scheme, as shown in the figure 2 is most widely used scheme [49]. The "10" and "20" signify that the actual distance between consecutive electrodes is 10% or 20% of the skull's overall front-to-back or right-to-left distance, respectively. Occipital electrodes are the most effective way to record the posterior dominant rhythm, which is the dominant frequency for sleep research and ranges from 8 to 13 Hz, according to a study by Malhotra *et al.* [50]. Hence we moved forward with extracting the single electrode signals from the MWT dataset [16] which accurately scores based on the Bern test criteria [51] for sleep staging and was created explicitly for classifying the wake-sleep transition zone and incorporating different labels for MSE (Microsleep Episodes). The research of Hertig *et al.* [16] developed the first constant and high-resolution MSE (Microsleep Episodes) grading standards, which verified with quantitative EEG analysis.

The EDA pipeline's approach is to execute the basic filtering and modulation required to standardize the data into a versatile format while keeping as much of the signal as feasible. To tackle the issue of class imbalance, we used three different methods:

1) Manual inspection of the plotted model showed that the 4th class had significantly fewer labels. Thus, the class labels were re-labeled, forming three training classes: Wake, MSE, and Sleep, and this approach improves the class ratio.
2) Adaption of Class Adaptive Loss Gradients weights which depends on the three classes' ratio.
3) Employing a custom method of Adaptive Resampling as explained in IV-B.

The original EEG data file of any subject consisting of long-duration signals is visualized in Fig. 4. The EEG data is stripped into 30s intervals to form the training corpus. A 50 Hz notch filter is then applied to the data because it is the frequency of the main line noise. A High-Pass fourth-order Butterworth Filter preceded the notching to remove long-term non-singularities in the signal. Figure 3 shows that the difference between signals before and after notching has higher magnitudes at the beginning of the signal. An infinite impulse response (IIR) filter was employed to extract the signal from the 4 to 14 Hz range. An ideal FIR Filter could be a good choice but is computationally expensive and susceptible to noise artifacts [52]. A biquid Chebyshev II filter is employed as the bandpass applicant. After that, the unnecessary artifacts were removed using power spectrum density analysis. After this, the signal was transformed into the formats needed to be fed into the neural network models. As mentioned earlier, the authors were interested in both temporal as well as spectral analysis of the data, and the required data conversions are explained in a detailed manner in IV-D and IV-C sections, respectively. After this, the data was split into train and test data with an 80:20 ratio. Further, the validation data was extracted from the newly formed training set with a 70:30 ratio. The data were stratified in all steps to have a uniform distribution of all the classes, and class weights were accordingly kept.

IV. Preliminary Background

As our model is designed for temporal and spectral events, it is intended on the following core principles: Robustness, Complementarity, Generalization, and over-fitting-aware training. The following section describes the necessary theory and proposes our methodology used in dataset creation after the pre-processing for the deep learning model.

A. *Dataset Selection Methodology*

The MWT test measures a person's ability to resist the desire to sleep in sleep-inducing circumstances, a process termed as "wake tendency." The MWT is the measure used in clinical settings that most closely simulates passive real-life situations, where MSEs (Microsleep Episodes) are crucial. Furthermore, the success of real-world drivers is highly correlated with the sleep identification latency in the MWT. As a consequence, MSEs in the MWT dataset can be classified with a high degree of temporal specificity The conditions set for the MWT test are similar to a scenario of how a driver sits in the car while driving. The MWT dataset[16] authors decided for this research work consists of 64 subject (aging from 30 to 70 years). Each file contains a MWT trial recording of a patient containing the labels of Wake-fullness, Episodes of drowsiness, Microsleep

Episodes and Microsleep episode candidates with labels in the ratio of 4:2:1:0.25 respectively(we had about 5000 labels for "Wakefullness" class). Original dataset contains the trials with multiple electrodes but we specifically extracted the data from one of the electrode for the analysis, details of which have been presented in the section III.

B. *Adaptive Resampling*

A typical neural network with various input types often integrates the characteristics learned from the component streams to generate a combined representation, which subsequently fulfills the classification function. Consequently, there is no approach to control the rate at which separate streams learn. Simultaneous generalization and overfitting of the model could resolve this issue. However, this is impossible due to the limited amount of "microsleep" data and excessive "wakefulness" data. As a result of an imbalanced input dataset, the model overfits even after applying regularisation to its training. To tackle this issue, we propose to adaptively oversample a number of features of "microsleep" and "deep sleep" classes by the following method: I. The average absolute difference between the data samples II. Analyzing the co-variance and correlation and III. Avoiding Redundancy while forming new data samples. Note that the presented method is not very complex and involves forming new samples by generating time-averaged signals, yet it is effective. This is because we incorporated step II. and step III. to account for preventing Redundancy in new samples.

Let us denote an epoch as E_i^c, where c defines the class of the epoch (W, M, S for Wake, Microsleep, and Sleep) and i denotes the serial number of the epoch from the individual corpus. Further, the individual amplitude for a particular time stamp can be defined as $E_i^c : e_t$. To measure class correlation, we found the maximum possible absolute difference between any two signals of the class corpus. Let us term this as a difference factor, denoted by D_c. For all combinations of i and j in the corpus:

$$D_{ij} = 1/N * \sum_{t=1}^{N} (E_i^c : e_t - E_j^c : e_t)$$

and $D_c = \max(D_{ij})$ for each class c. Note that N is the normalization factor equal to 6000, i.e., the number of samples in an epoch (200Hz and 30s epoch length).

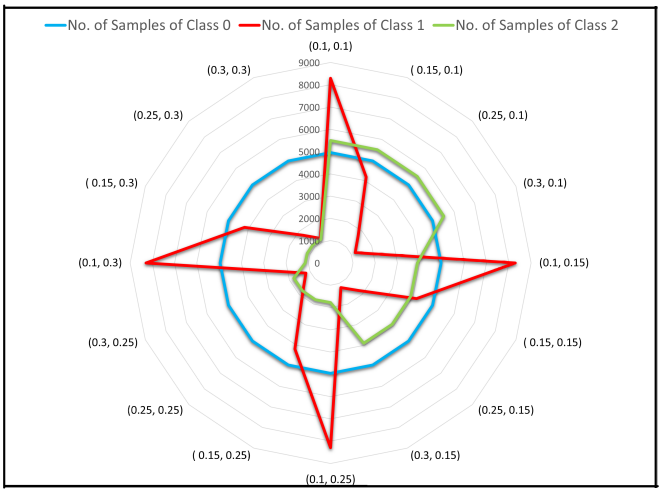

Fig. 5: Cross Power Coherence Threshold Optimization Curve for Adaptive Resampling. The Blue contour is for Wakefulness, the Red contour is for Microsleep, and Green contour is for Deep Sleep.

For generating new signals from the class corpus, we time averaged any pair of signals whose normalized absolute difference strength is greater than $(D_c * Th_c)$, where Th is the class threshold necessary to regulate the number of new samples formed. This ensures the formation of functionally new signals that improve the testing statistics (refer section VI-C for results). For all combinations of i and j in the corpus we define modified E^c as:

$$Insert = \begin{cases} E_i^c, E_j^c and (E_i^c + E_j^c)/2, & \text{if } D_{ij} \geq D_c * Th_c \\ E_i^c and E_j^c, & \text{otherwise} \end{cases}$$

The choice of Th_c is decided by plotting the Threshold Optimization Curve, as shown in Fig. 5. The values in the brackets are the possible threshold values for Microsleep (Th_M) and Deep Sleep (Th_S), which are used to find the optimal values. In the figure, we can see the Blue contour is for Wakefulness, the Red contour is for Microsleep, and the Green contour is for Deep Sleep. Here, the Labels for Wake fullness, Microsleep, Microsleep Episodes, and Deep Sleep are 0, 1, 2, and 3, respectively. We take Th_W as 1 and find other thresholds in reference to it. As we can see in the figure, all the three class contours approach each other when $Th_M = Th_S = 0.15$. Applying this resampling improved the overall robustness of the model as we got nearly equal(4957, 4180 and 3899 samples of each classes respectively) number of unique samples for each classes. The original ratio from the dataset was thus improved from 4:2:1 to 1.18:1:0.932, thus proving our proposed resampling method's impact.

C. *Spectral Analysis*

Time series data, like EEG, can be represented as time-power-frequency images. Using Fourier-related transforms, the Short-time Fourier transform (STFT) is used to extract the sinusoidal frequency and phase content of short sections of a time-varying signal [53]. Engineers use this kind of visual extract signals resonant in the space where the signal was recorded. Theoretically, computing STFTs entails dividing a larger temporal signal into equal-length segments and performing the Fourier transform independently on each shorter segment. In each shorter section, the Fourier spectrum is revealed. The spectral estimate is based on Fourier analysis, which presupposes infinite signals, continuity, periodicity, and stationarity in the data. However, none of these assumptions are frequently met in EEG data; therefore, the spectrogram is a severely biased estimate in practice. We convolve a periodic

Algorithm 1 Proposed Algorithm to develop STWT $\Phi(signal)$

Require: $n \leq 6000$
Require: $t \leq 30$

$windowSize \leftarrow winSize$
$winSlide \leftarrow winSize/2$ ▷ To ensure 50% overlap
$k \leftarrow \max n + winSize$ ▷ Padding
$iter \leftarrow 2 + 2 * (\max n - winSize)//winSize$
$tempLim \leftarrow 0$

if $k \geq winSize/2$ **then**
  $x(k) \leftarrow signal$
else
  if $k \leq (\max n - winSize/2)$ **then**
    $x(k) \leftarrow signal$
  else
    $x(k) \leftarrow 0$
  end if
end if

while $iter \neq 0$ **do**
  $f, t, Zxx \leftarrow STFT(x(tempLim : tempLim + winSize))$
  $power \leftarrow |Zxx|$
  $\Phi(\ iter) \leftarrow \Sigma power$ ▷ axis = time
  $iter \leftarrow iter - 1$
  $tempLim = tempLim + winSlide$
end while

Hann window taper over the raw signal, as a strategy to reduce the spectrogram's bias before performing spectral estimation.

$$\omega(n) = 0.5(1 - \cos(2 * \pi * n/N - 1)), 0 \leq n \leq N - 1$$

Here, $\omega(n)$ is the Hann The window, with the maximum value normalized to one and N is the number of points in the output window. The STFT spectrogram also has the significant flaw of producing estimates with considerable variance across all frequencies.

*1) **Proposed Short Time Welch's Transform Methodology**:* Suppose we conduct an FFT over the signal in such a real and practical situation of microsleep detection. In that case, we are missing that it is a sample of a dynamic system; hence, much of its content is merely unwanted noise. Welch's transform aims to precisely get rid of the stochastic noise by the method of averaging. It divides the original signal into many fragments and averages their spectra [54]. The Welch's PSD (Power Spectrum Density) method is a popular non-parametric improvement on the standard periodogram spectrum estimating methodology, which utilizes FFT to reduce the non-stationary noise in exchange for reducing the frequency resolution of the estimated power spectra [55]. Another major benefit of this strategy is that it reduces the number of computations and the amount of core storage required. This approach entails sectioning a signal, modifying the periodograms of each segment, and averaging the changed periodograms.

Keeping these factors in mind, we formalize an algorithm to generate Overlapping Short-time-Welch's transform (STWT) that adapts to cater to the needs of biomedical signal processing, which generally contains lots of random noise. The signal is split up into overlapping segments (say K data segments) of length L, overlapping by O points. For our case, considering $O = L/2$, we can say that we have a 50% overlap. Assuming that

$$x(t : t_d), 0 \leq t \leq t_d \leq 30 x(n : n_d), 0 \leq n \leq n_d \leq 6000$$

denotes an epoch starting from $time = t$ to $time = t_d$, and $x(n)$ being the same epoch in the sampled and discrete state, we make use of STFT to derive the formula for STWT. We first state the formal algorithm for STFT:

$$X(n, w) = \sum_{m->-\infty}^{\infty} x(m) * w(n - m) * \exp(-jwn)$$

where $x(m)w(n - m)$ is a short time part of the input signal $x(n)$. The complete algorithm for STWT is presented in the algorithm 1, where we ensure a 50% overlap with padding to creating these small windows of signal. These individual windows of intervals are processed through STFT and averaged out to mimic the Welche's method affectively. The spectrum thus formed is presented is a 3-D representation of the dynamics of the time-frequency-power of the spectra. A Hanning Window gives the data in the center of the set greater weight than the data on the edges, resulting in information loss. Individual data sets are frequently overlapped in time to reduce this loss. Figure 6, clearly shows the Spectro-Temporal Power Spectrograms for their corresponding raw signals and establishes that STWT extracts required features to secern easily the two classes along the entire 30s epoch.

D. Temporal Analysis

Temporal Analysis refers to the analysis done over the raw variant of the desired signal. This analysis is practical and explainable in terms of deep learning models, as we can visualize the steps. However, it suffers from 2 significant drawbacks: I. The number of parameters increases significantly compared to Spectral Analysis, and II. Microsleep is represented in terms of frequency. Thus, it is impractical to use a temporal analysis for the same. However, we have used two modes of data for temporal analysis purposes - **raw mode** - which consisted of the original signal epochs (After pre-processing), and **norm mode** - which consisted of the normalized version of the raw signals. Transforming the values of numeric columns in a dataset to the same scale without distorting the ranges of values is referred to as normalization. For our case, the Frobenius normalization was preferred over standardization and normalization because these methods had higher amplitude values and linear scaling of the raw signals. The log Normalization method was also assessed, but experimental results only favor Frobenius Normalization.

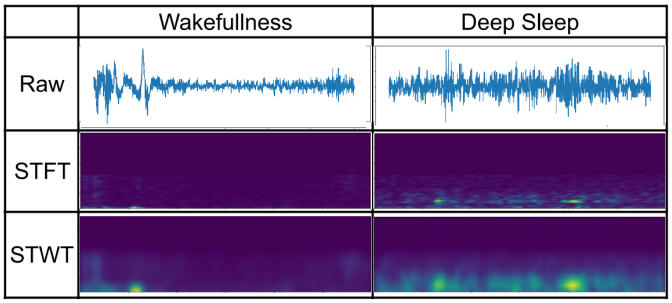

Fig. 6: Comparative results of STFT and the proposed STWT methods. The first row represents the raw images, the second row represents the spectrogram formed using STFT and the third row represents the spectrogram formed using STWT. The x-axis denotes time stamps.

Different types of experimented normalization are defined below:

$$ForbeniusNorm \leftarrow x(t)/(\Sigma|x(t)|^2)^{(0.5)}$$

$$LogNorm \leftarrow \log x(t)$$

V. Proposed Training Framework

As the analysis involves both temporal and spectral analysis, the model has been developed using a unique part that captures its details, followed by specific reusable for both the respective counterparts. Figure 7 shows the overall implementation details of the spectral and the temporal analysis, and figure 8 embarks the legend to understand the models. We are using the last half of the model for both temporal and spectral signals. Initially, both networks aim to make the input become 2-D data. The Temporal Model is passed through multiple 1-D filter banks with different convolutional parameters. Later, both networks include a Multi-Headed Attention layer followed by 3 Multi-Fusion extractors. Finally, the model gives a softmax output with the probabilities of different sleep modes. The following section includes the details of the two networks: I. The Spectrum Image passes through Spectral Model, followed by the Multi-Headed Attention block, followed by Multi-Fusion Extractor, and finally, the output layer. II. Raw signals first go through Temporal Model, followed by the Multi-Headed Attention block, followed by Multi-Fusion Extractor, and finally, the output layer.

A. *Spectral Model : Two Layered Net*

When fed to this unit, the original 30-second signal epoch has already been transformed into a power spectrum. To improve the model's interpretability, some prior studies [56] recommended transforming the spectrum to a logarithm scale and producing log-power spectrum images. The output image is filtered using a spectral filter bank for frequency smoothing and dimension reduction. We refrained from using a standard upper triangular filter and preferred learning the filter bank precisely for the task. According to previous works [56], a discriminatively learned filter bank using an Artificial Neural Network is more capable of automated sleep staging than a standard one. The learned filter bank should accentuate the sub-bands that are more critical for the job while attenuating the sub-bands that are less significant. As a result, for preprocessing, we employ a filter bank that has been pre-trained using a Deep Neural Network.

A Residual Squeeze and Excitation block [57] follows a simple 2-D Conv layer with a leaky-Relu activation, which seeks to re-calibrate the features learned by the preceding Conv block in order to improve its performance. This SE block, in particular, models feature inter-dependencies and adaptively picks the most discriminative features. It also assists in formulating a context-aware technique to let the network's lower layers use additional relevant information beyond its immediate receptive area. A global average pooling (GAP) layer compresses a 2-dimensional data kernel to a 1-dimensional point variable that is the average of all those data points, squeezing the block's global spatial information. After then, two completely linked layers are used to utilize the synchronized data. A ReLU activation function limits the dimensionality in the first layer, whereas a smoothing sigmoid activation function expands the second layer's dimensionality. The modality fusion component is then adjusted by multiplying the output with the original input:

$$\dot{X} = X * GAP(\Psi(\Psi(X))))$$

where, Ψ denotes the dense layers, X is the input and $\dot{X}$ is the output and GAP is global average pooling.

Following this layer, the second network set was established, involving multiple dilated 1-D causal convolutional blocks. Causal convolutions can encode and capture the location information of input data and their temporal relationships. Compared to RNNs, these causal convolutions offer the benefit of quick and parallel processing, considerably saving model training time. This was necessary because the next part involved a multi-headed attention unit that works with positional encoding.

B. *Temporal Model : Raw2Img*

We employ two CNNs with small and large filter sizes in the initial layers to extract time-invariant properties from raw single-channel 30-s EEG epochs. This ensured a balanced trade-off between temporal and spectral accuracy in their feature extraction [58]. The bigger filter captures frequency information (i.e., frequency components), whereas the smaller filter captures temporal information (when particular EEG patterns arise). First, each of the two branches is smoothed in spectral dimension using a trainable filterbank block of K filters and reduced frequency bin count from F to K. The kernel sizes S are chosen based on the sampling rate of the EEG signals and the goal of exploring different frequency bands. Furthermore, different frequency ranges are associated with distinct sleep phases [59], making it necessary to address diverse frequency bands to optimize the retrieved characteristics. As a result, we employ kernel widths to capture time-step ranges and address

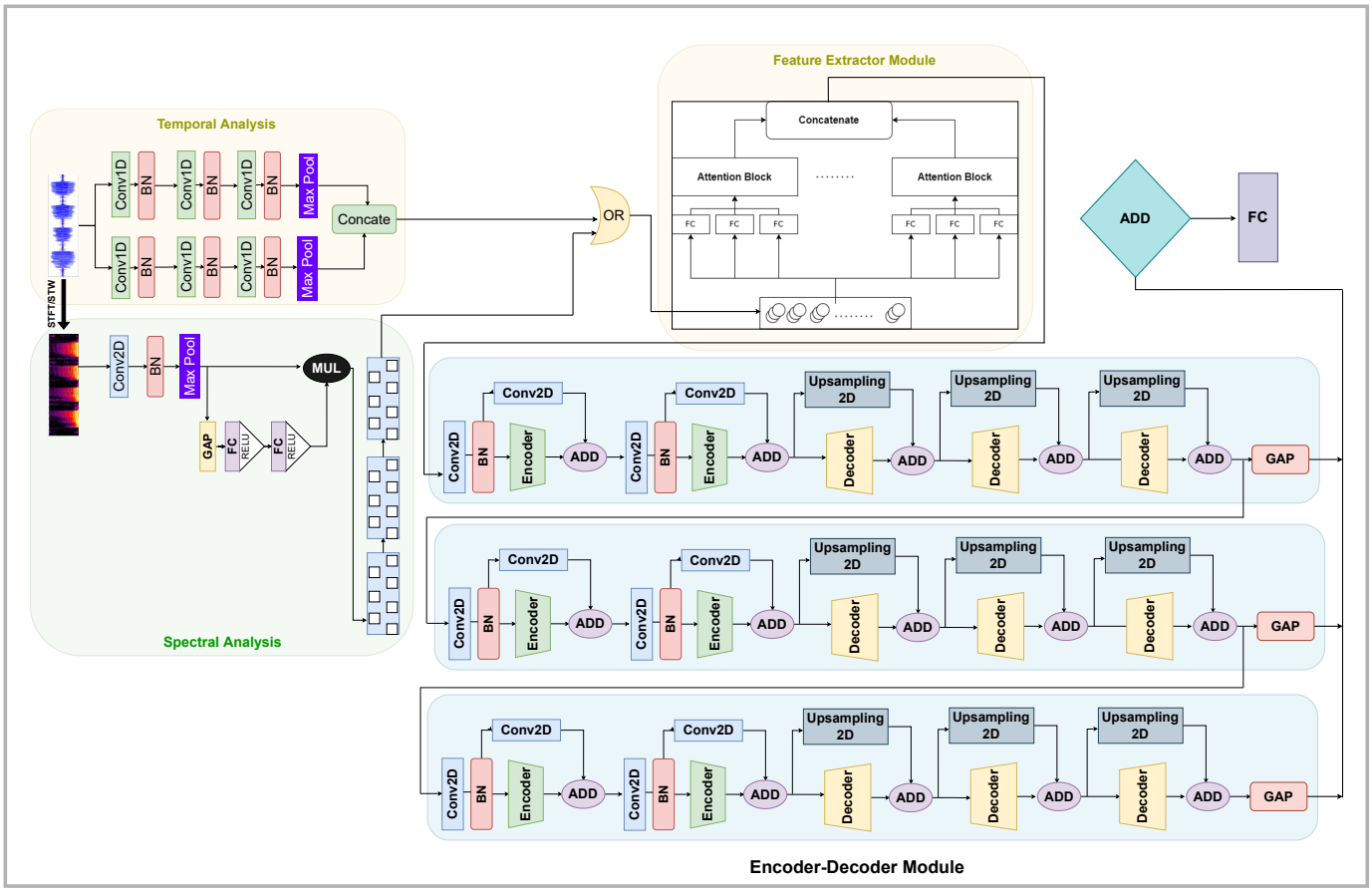

Fig. 7: The overall end-to-end architecture of our proposed classification framework

characteristics from various sleep-related frequency bands. The non-stationary nature of EEG data necessitates the exploration of many types of characteristics, which necessitates such a combination of features. In the network, each of the two branches consisted of three 1-D convolutional filters, which align closely with the clinical signal processing techniques and uses an activation function of the Gaussian Error Linear Unit. Multiple max-pooling layers, batch normalization layer, and 50% dropout rates were deployed to avoid overfitting and reduce the latency. Both the branches were concatenated along the filterbank axis, i.e., the number of filters was added while mainly reducing the overall signal length. The Gaussian Error Linear Unit (GELU) activation function is also used. Other activations, such as Leaky-ReLU, which transfer negative values, may have been better than GELU since they let strong negative activations negatively affect the overall result of other activations feeding the subsequent layers and leading to undesirable outcomes. GELU, on the other hand, appears to have better control over the consequences of these negative activations.

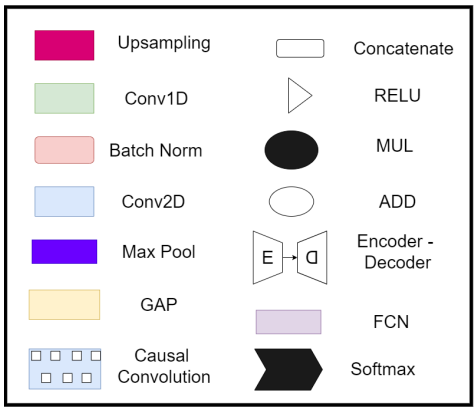

Fig. 8: Legend to understand the overall neural network model.

C. *Attribute extraction unit - Common block 1*

Softmax threshold is used to reduce superfluous information from the previous blocks, following a multi-headed attention link to filter and emphasize critical information. Such a single attention block can be described by [60]:

$$\Psi(K, Q, V) = sigmoid(K.Q'/\sqrt{d})) * V$$

where d is the length of the signal, and K, Q and V are the input vectors to the attention block. In our case, we are using self-attention, which is a slight modification of the above definition and is defined as follows:

$$\dot{\Psi}(X, X, X) = sigmoid(X.X'/\sqrt{d})) * X$$

A Multi-headed attention block is an outcome of the concatenation of multiple such units. This technique promotes

the model's self-awareness by increasing the model's capacity to focus on diverse places since each head's encoding is aware of the encodings of the other heads. The model's capacity to learn temporal dependencies is improved due to this. Additionally, segregating the input features widens the representation subspaces. As a result, the attention poundage created for individual fragments reflects the relevance of each subspace. We yield improved classification accuracy by forming richer representations by grouping the above-mentioned representations.

D. *ENDEC - Common Block 2*

This part of the model dealt with developing a series of encode-decoder(ENDEC) networks acting as feature calibrators whose outputs are passed through Global Average Pooling (GAP) layers, which finally generate the output. The following equation summarizes the architecture, considering x as the input matrix and $ACF = Encoder(x) + Decoder(x)$:

$$Output = GAP_1(ACF_1(x)) + GAP_2(ACF_2(ACF_1(x))) + GAP_3(ACF_3(ACF_2(ACF_1(x))))$$

The individual encoders and decoders had skip connections having a 2-D Conv Layer to incorporate attention-based feature calibration, inspired by the work by Xue Jiang [61].

$$Encoder = \rho(\aleph(\S(\aleph(\S(x))))) + \urcorner(x)$$

$$Decoder = \rho^{\intercal}(\aleph(\urcorner^{\intercal}(\aleph(\urcorner^{\intercal}(x))))) + \urcorner(\rho^{\intercal}(x))$$

where, $\S$, $\urcorner$, ρ and $\aleph$ stands for Separable Convolutions, Normal Convolution, Pooling layers and Batch Norm layers. Their corresponding transpose layers have also been used in the same equation.

The encoders are primarily built using Separable Convolutions which minimises the amount of parameters in a standard convolution. If the network is already tiny, the approach may end up with only a few parameters, preventing it from training successfully. However, as in our case, the technique tends to improve the overall efficiency without sacrificing the efficacy, making it the perfect choice for reducing latency. Everywhere in this module, ReLU activation was used. GELU should outperform ReLU because ReLU transforms any negative weights to zeros, preventing the module from utilising them. In out module, however, we use ReLU since it aims to reduce exploding/vanishing gradients while simultaneously speeding up and simplifying computations [62].

E. *Output Generating Module*

The final layer addresses a Dense layer followed by a softmax output to get the probabilistic overview of the predicted output. Kernel Regularizers are also used, which penalize the layer parameters during optimization.

VI. Results and Discussion

In Figure 10, we show the comparative results of our model with the state-of-the-art model [63], which is developed using the same dataset type. These methods include the ones that utilize window-based CNN and LSTM methods. The Cohen's Kappa of the expert algorithms and the inter-rater agreement was suitable for MSEs and wakefulness (nearly 0.7) but was negligibly low for Deep Sleep ($\leq$0.1). It can clearly show that our proposed method delivers a significant improvement in the results. Among the methods we have presented, the temporal analysis with the raw signals has shown the best results. Our model produced similar results without reducing MSEc and ED performance, which is the most challenging sleep stage to interpret. This indicates that the robustness of our method is restrained from being skewed towards the majority of the sleep stages. In addition, the performance of multiple models on the validation set was evaluated using the proposed proposition to extract individual Cohen's Kappa values. It measures inter-rater reliability and, thus, considerably more robust indicator than other matrices. The cohen-Kappa value of a distribution can be mathematically formulated as follows:

$$\kappa = (P_{r_a} - P_{r_e})/(1 - P_{r_e})$$

where P_{ra} is the relative observed agreement among raters,

Algorithm 2 Proposed method to find individual kappa scores

Require: $pred \leftarrow ModelPred(X)$
Require: $true \leftarrow Y$
$classes \leftarrow 3$ ▷ We have three classes

for class in classes **do**
  if $pred == class$ **then**
    $pred_n = 1$
  else
    $pred_n = 0$
  end if
  if $true == class$ **then**
    $true_n = 1$
  else
    $true_n = 0$
  end if
  $\kappa(class) \leftarrow CohenKappa(true_n, pred_n)$
end for

and P_{re} is the hypothetical probability of chance agreement. Using the overall κ coefficient, it can be seen that the agreement between sleep specialists and our model is very high (93, 76, 81, 84 respectively for the raw, norm, stft and stw based analysis). Generally, the kappa score shows the overall performance , but to calculate the individual kappa scores for each classes, we had to slightly modify our calculations by vectorizing individual classes, as shown in algorithm mentioned in algorithm 2. A n-sized vector was created for each prediction, and this was masked with another n-sized vector where the nth position was the position of the class we are interested in. Kappa scores on these vectors can be represented as the individual kappa scores of the classes.

A. *STWT vs STFT*

Figure 11 clearly shows that our proposed STWT method performs better than the traditional STFT method. The features

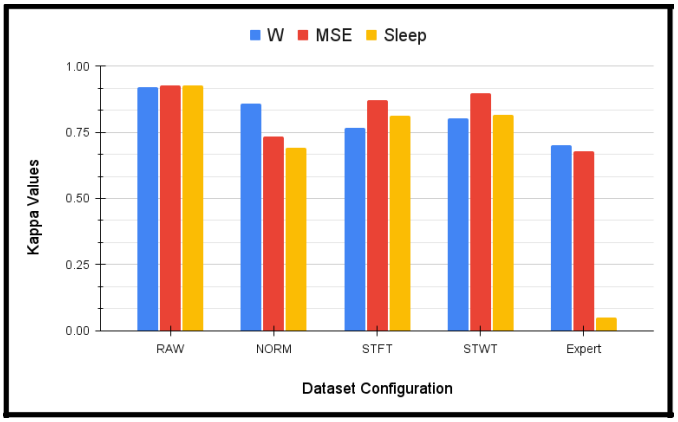

Fig. 10: Bar Chart for comparing our results with the previous state-of-the-art model using the same dataset type.

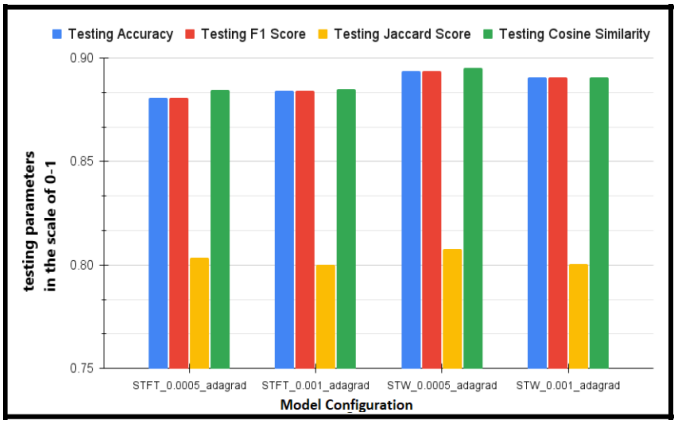

Fig. 11: Quantitative comparison between the STWT and STFT for different testing parameter. The x-axis showcases the configuration of different models used and the Y-axis quantifies testing parameters used.

in the image showcase the improvement of about 1-2% in the testing set for all different evaluation matrices. The f-1 scores for the STFT-based models were around 88%, whereas, for STWT-based models, the scores were between 89-90%. In the case of training and validation datasets, the improvement was nearly about 3% and 2%, respectively. Not only this, as shown in figure 6, the STWT method increases the robustness of the methodology by providing visually explainable features and yet outputs smaller and richer images. Although we see very less improvement in the accuracy, the proof of concept of our method has very well been established. Further, if higher frequency signals were to be analyzed using our method, the results would have been much richer because most of the noise and non-stationarities lie in the upper band of frequencies, which our method can positively counter.

B. Parameters

All four modes were trained recursively using four parameters: a learning rate of 0.0005 and 0.001 and optimizers of Adagrad and RMSProp. The Spectral Models gave better results with Adagrad Optimizers, while the Temporal models gave much improved results with RMSProp. Moreover, the shown statistics for the Spectral analysis are for the unsampled method, and the ones for the temporal analysis are for the resampled method. The reason for the same is explained in section VI-C. Further, in all the cases, a lower initial learning rate gave better results since the model is quite large, and many parameters would die away if the learning rate were large. To further improve the accuracy, we also employed an Exponential Decay learning rate scheduler and early stopping with a 50% patience. The two charts in figure 9 show the final Training, Validation, and Testing Normalized Accuracy and Cosine Similarity of the different parameters for the Spectral and Temporal Models, respectively. One interesting trend we observe is that Spectral analysis gives poor testing results, whereas temporal analysis gives an overall better result for all four models. Even though Norm_0.001_RMSProp seems to perform poorly, its parameters are still at about 90% !

The cosine similarity is advantageous to measure testing in our case because even if the two similar signal labels are far apart by the Euclidean distance, they may still be oriented

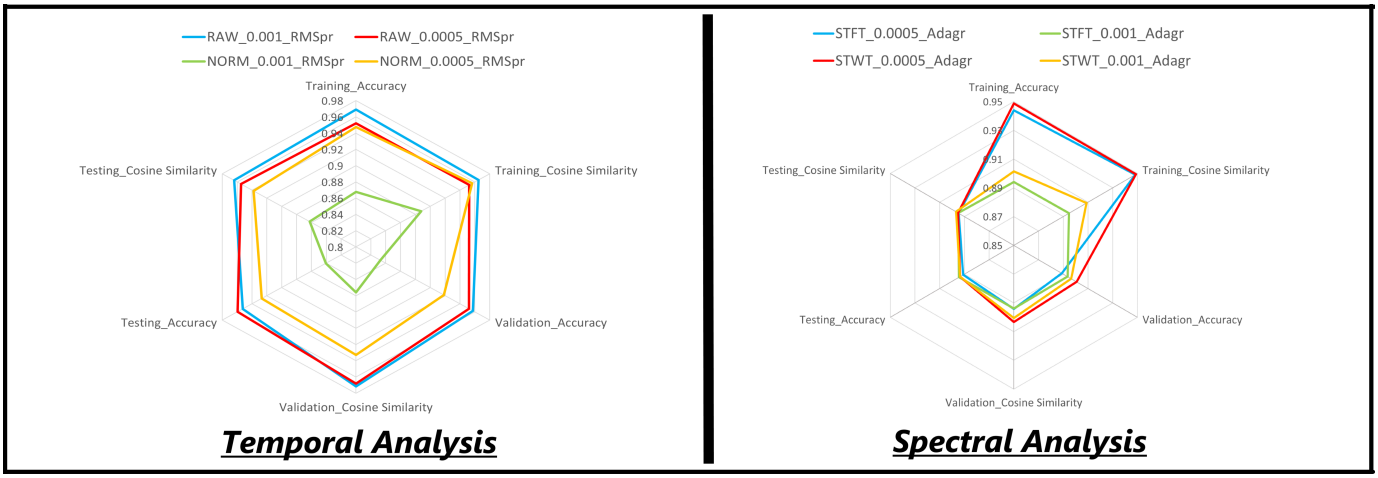

Fig. 9: Accuracy and Cosine Similarity charts for the training, testing and validation sets of the Temporal and Spectral Analysis.

Dataset Config	Model Config	Training Matrices				Validation Matrices				Testing Matrices			
		Accuracy		Cosine Similarity		Accuracy		Cosine Similarity		Accuracy		Cosine Similarity	
		Unsampled	Resampled	Unsampled	Resampled	Unsampled	Resampled	Unsampled	Resampled	Unsampled	Resampled	Unsampled	Resampled
RAW	0.001_rmsprop	0.91	0.97	0.92	0.96	0.90	0.96	0.90	0.97	0.90	0.95	0.90	0.96
	0.0005_rmsprop	0.95	0.95	0.95	0.95	0.90	0.95	0.91	0.97	0.90	0.96	0.91	0.95
NORM	0.001_rmsprop	0.96	0.87	0.96	0.89	0.90	0.83	0.90	0.86	0.90	0.84	0.90	0.86
	0.0005_rmsprop	0.91	0.95	0.92	0.96	0.89	0.92	0.89	0.93	0.89	0.93	0.90	0.94

Fig. 12: Table of results(rounded off) showing the statistics of different models with resampled and unsampled inputs. We have compared 2 models with config 0.001_rmsprop and 0.0005_rmsprop learning rates respectively.

closer together. Such a case can qualitatively occur if we re-define the region of Micrsoleep during the transition of Deep Sleep and Wakefulness, thus having an equal positive correlation with both classes.

C. *Resampled vs Unsampled*

Our employed adaptive sampling technique resulted in a significant improvement. The dataset size increased in such a fashion as to over-sample the under-sampled classes and yet avoid repetition by maintaining the diversity of the dataset. As shown in figure 12, we observe a noticeable increase in the validation matrices of the model after sampling. The statistics show a 4-5% increment in the cosine testing accuracy while a 6-7% and 1-4% increment in the validation and training accuracy. The resultant graphs for RAW (Fig - 13) and NORM (Fig - 14) dataset configs show that the resampled models outperformed the unsampled models in all the cases.

The Spectral Mode, however, suffered a considerable dip in accuracy after resampling with about 2-6%, 1-5%, and 3-10% loss in testing, validation and training matrices, with an exception from the STFT model having a learning rate of 0.0005 and Adagrad optimizer, which showed a positive trend. One of the possible reasons for this drop could be that our resampling paradigm involves adaptive averaging of 2 signals having a large difference in their average powers to produce new signals. One can reason that the time-averaging signals do not affect the spectral front. Thus, our method duplicated the parent signals, thus reducing the overall efficiency of the model. This theory was also backed by low overall Hamming loss (which is defined as $\dot{N}/(N+\dot{N})$, where N is the number of correct labels) of these models, even after resampling. This result points us to the fact that the model is overfitting to predict only the label having the highest occurrence in the training set, i.e., Wakefulness.

VII. Conclusion

In this paper, we proposed a Welch-Transform-based Enhanced Spectro-Temporal Analysis for Cognitive Microsleep Detection using a Single Electrode EEG. We formulate a spectro-temporal classification problem and propose two distinct Neural Network techniques for detecting microsleep during automobile transportation utilizing maintenance of wake-fullness(MWT) test data. We also highlight the importance of choosing the correct dataset and the significance of appropriate data cleaning and pre-processing as part of an end-to-end architecture. In particular, we develop a hierarchical Convolutional Encoder-Decoder network to address this challenge with multiple skip connections and attention units. The network is trained from start to finish using dynamic folding and expanding of the input sequence at various levels of the network structure.

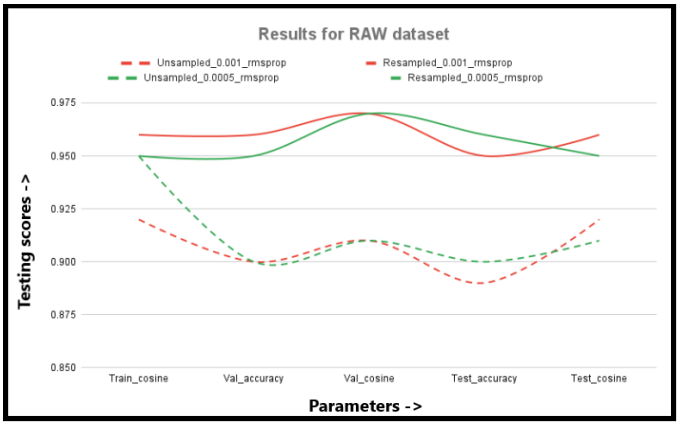

Fig. 13: Graphical representation of the resultant matrices for Raw dataset processing case.

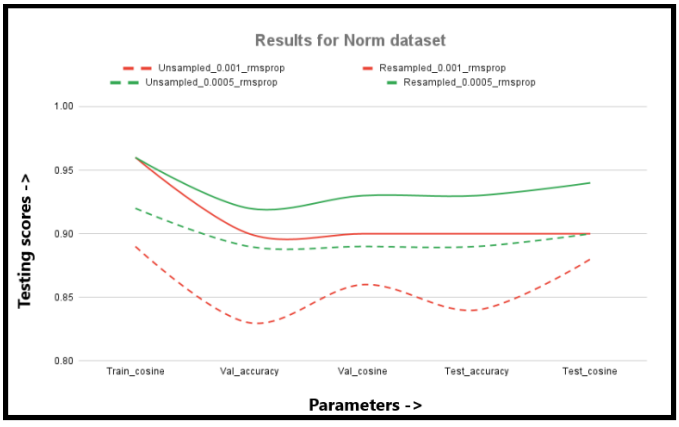

Fig. 14: Graphical representation of the resultant matrices for Norm dataset processing case.

Our results show that cosine similarity varies with different parameters and demonstrate the choice of cosine similarity as the best evaluation matrix for classifying transitional parameters such as microsleep. Our proposed methods surpass prior work and have strong baselines for comparison, achieving state-of-the-art performance throughout the full MWT dataset [16]. Our proposed adaptive resampling methodologies are also shown to reduce the influence of the class-imbalance problem and improve performance, particularly for the case of our model's performance in predicting Deep-Sleep, which is more difficult to score than other sleep stages. The primary goal of this study is to show that, regardless of the complexity or kind of architecture employed, the use of the STWT for spectral analysis and resampled inputs for temporal analysis yield better outcomes than traditional approaches. Our current future work aims to optimize our end-to-end architecture for real-time processing capability for future clinical evaluation and deployment.

VIII. Acknowledgement

Hussain would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) - Grants Ref. EP/M026981/1, EP/T021063/1, EP/T024917/1.

References

[1] D. Yeo, J. Lee, W. Kim, M. Kim, K. Cho, A. Ataya, and S. Kim, "A hand-over notification system of vehicle driving control according to driver's condition based on cognitive load in autonomous driving situation," *Proceedings of the HCI Korea, Seogwiposi, Korea*, pp. 13–16, 2019.

[2] B. Mok, M. Johns, K. J. Lee, D. Miller, D. Sirkin, P. Ive, and W. Ju, "Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles," in *2015 IEEE 18th international conference on intelligent transportation systems*. IEEE, 2015, pp. 2458–2464.

[3] I. D. Luce, "Your brain might be taking tiny naps throughout the day, and it can lead to disaster if left unchecked." [Online]. Available: https://www.businessinsider.com/what-is-microsleep-signs-youre-sleep-deprived-2019-6

[4] "What to know about microsleep: Dangers, causes, and prevention." [Online]. Available: https://www.webmd.com/sleep-disorders/what-to-know-microsleep

[5] D. Welle, "Asleep at the wheel: Driving while drowsy can be deadly," 10 2019. [Online]. Available: https://www.dw.com/en/asleep-at-the-wheel-driving-while-drowsy-can-be-deadly/a-50653206

[6] I. D. Luce, "Drowsiness reports while driving a car," Jun 2019. [Online]. Available: https://www.businessinsider.in/

[7] G. R. Poudel, C. R. Innes, P. J. Bones, R. Watts, and R. D. Jones, "Losing the struggle to stay awake: Divergent thalamic and cortical activity during microsleeps," *Human Brain Mapping*, vol. 35, no. 1, p. 257–269, 2012.

[8] A. Paul, L. Boyle, J. Tippin, and M. Rizzo, "Variability of driving performance during microsleeps," 01 2005.

[9] B. B. Hazarika, D. Gupta, and B. Kumar, "Eeg signal classification using a novel universum-based twin parametric-margin support vector machine," *Cognitive Computation*, pp. 1–16, 2023.

[10] V. Kohli, U. Tripathi, V. Chamola, B. K. Rout, and S. S. Kanhere, "A review on virtual reality and augmented reality use-cases of brain computer interface based applications for smart cities," *Microprocessors and Microsystems*, vol. 88, p. 104392, 2022.

[11] A. Chougule, J. Shah, V. Chamola, and S. Kanhere, "Enabling safe its: Eeg-based microsleep detection in vanets," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[12] A. Balaji, U. Tripathi, V. Chamola, A. Benslimane, and M. Guizani, "Toward safer vehicular transit: implementing deep learning on single channel eeg systems for microsleep detection," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[13] D. Leger, "The cost of sleep-related accidents: a report for the national commission on sleep disorders research," *Sleep*, vol. 17, no. 1, pp. 84–93, 1994.

[14] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.

[15] B. C. Committee, "2020 international bci competition," April 2020. [Online]. Available: https://doi.org/10.17605/OSF.IO/PQ7VB

[16] H.-G. Anneke, S. Jelena, M. Alexander, A. Peter, M. Johannes, and S. D. R., "Maintenance of wakefulness test (mwt) recordings," Sep. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3251716

[17] D. Alvarez-Estevez and R. Rijsman, "Haaglanden medisch centrum sleep staging database," Mar 2022. [Online]. Available: https://doi.org/10.13026/t79q-fr32

[18] M. Ghassemi, B. Moody, L.-W. Lehman, C. Song, Q. Li, H. Sun, B. Westover, and G. Clifford, "You snooze, you win: The physionet/computing in cardiology challenge 2018," *2018 Computing in Cardiology Conference (CinC) Computing in Cardiology Conference (CinC)*, 2018.

[19] G. Zhao, Y. Li, and Q. Xu, "From emotion ai to cognitive ai," *International Journal of Network Dynamics and Intelligence*, vol. 1, no. 1, pp. 65–72, 2022.

[20] B. Li, Z. Xu, N. Hong, and A. Hussain, "A bibliometric study and science mapping research of intelligent decision," *Cognitive Computation*, vol. 14, no. 3, pp. 989–1008, 2022.

[21] Y. Susanto, E. Cambria, B. C. Ng, and A. Hussain, "Ten years of sentic computing," *Cognitive Computation*, vol. 14, no. 1, pp. 5–23, 2022.

[22] A. V. Desai and M. A. Haque, "Vigilance monitoring for operator safety: A simulation study on highway driving," *Journal of safety research*, vol. 37, no. 2, pp. 139–147, 2006.

[23] Z. Zhang, S. Han, H. Yi, F. Duan, F. Kang, Z. Sun, J. Solé-Casals, and C. F. Caiafa, "A brain-controlled vehicle system based on steady state visual evoked potentials," *Cognitive Computation*, pp. 1–17, 2022.

[24] M. L. Jackson, G. A. Kennedy, C. Clarke, M. Gullo, P. Swann, L. A. Downey, A. C. Hayley, R. J. Pierce, and M. E. Howard, "The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness," *Accident Analysis & Prevention*, vol. 87, pp. 127–133, 2016.

[25] X. Li, Y. Yan, J. Soraghan, Z. Wang, and J. Ren, "A music cognition–guided framework for multi-pitch estimation," *Cognitive computation*, pp. 1–13, 2022.

[26] A. Eskandarian and A. Mortazavi, "Evaluation of a smart algorithm for commercial vehicle driver drowsiness detection," in *2007 IEEE intelligent vehicles symposium*. IEEE, 2007, pp. 553–559.

[27] S. Asadzadeh, T. Y. Rezaii, S. Beheshti, and S. Meshgini, "Accurate emotion recognition utilizing extracted eeg sources as graph neural network nodes," *Cognitive Computation*, vol. 15, no. 1, pp. 176–189, 2023.

[28] C. Ieracitano, N. Mammone, A. Bramanti, S. Marino, A. Hussain, and F. C. Morabito, "A time-frequency based machine learning system for brain states classification via eeg signal processing," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[29] C. Ieracitano, N. Mammone, A. Hussain, and F. C. Morabito, "A novel explainable machine learning approach for eeg-based brain-computer interface systems," *Neural Computing and Applications*, vol. 34, no. 14, pp. 11 347–11 360, 2022.

[30] T. Gao, J. Zhou, H. Wang, L. Tao, and H. K. Kwan, "Attention-guided generative adversarial network for whisper to normal speech conversion," *arXiv preprint arXiv:2111.01342*, 2021.

[31] H. Ma, F. Wu, Y. Guan, L. Xu, J. Liu, and L. Tian, "Brainnet with connectivity attention for individualized predictions based on multi-facet connections extracted from resting-state fmri data," *Cognitive Computation*, pp. 1–15, 2023.

[32] M. S. Aldrich, "Automobile accidents in patients with sleep disorders," *Sleep*, vol. 12, no. 6, pp. 487–494, 1989.

[33] S. Garbarino, N. Lino, M. Beelke, F. D. Carli, and F. Ferrillo, "The contributing role of sleepiness in highway vehicle accidents," *Sleep*, vol. 24, no. 2, pp. 201–206, 2001.

[34] A. S. Houssaini, M. A. Sabri, H. Qjidaa, and A. Aarab, "Real-time driver's hypovigilance detection using facial landmarks," *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems, WITS 2019*, 4 2019.

[35] X. Li, S. Lai, and X. Qian, "Dbcface: Towards pure convolutional neural

network face detection," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
[36] Ş. Karahan and Y. S. Akgül, "Eye detection by using deep learning," in *2016 24th Signal Processing and Communication Application Conference (SIU)*. IEEE, 2016, pp. 2145–2148.
[37] C. Gou, Y. Wu, K. Wang, K. Wang, F.-Y. Wang, and Q. Ji, "A joint cascaded framework for simultaneous eye detection and eye state estimation," *Pattern Recognition*, vol. 67, pp. 23–31, 2017.
[38] C. Ieracitano, F. C. Morabito, A. Hussain, and N. Mammone, "A hybrid-domain deep learning-based bci for discriminating hand motion planning from eeg sources," *International journal of neural systems*, vol. 31, no. 09, p. 2150038, 2021.
[39] M. S. Kalas and B. Momin, "Stress detection and reduction using eeg signals," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016, pp. 471–475.
[40] J. Perrier, S. Jongen, E. Vuurman, M. Bocca, J. Ramaekers, and A. Vermeeren, "Driving performance and eeg fluctuations during on-the-road driving following sleep deprivation," *Biological psychology*, vol. 121, pp. 1–11, 2016.
[41] Y.-S. Kweon, H.-G. Kwak, G.-H. Shin, and M. Lee, "Automatic microsleep detection under car-driving simulation environment using night-sleep eeg," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2021, pp. 1–6.
[42] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation. arxiv 2015," *arXiv preprint arXiv:1505.04597*, 2019.
[43] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of eeg sleep stage scoring," in *2017 IEEE 27th international workshop on machine learning for signal processing (MLSP)*. IEEE, 2017, pp. 1–6.
[44] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Seqsleepnet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 3, pp. 400–410, 2019.
[45] S. Mousavi, F. Afghah, and U. R. Acharya, "Sleepeegnet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PloS one*, vol. 14, no. 5, p. e0216456, 2019.
[46] X. Jiang, "Mrnet: a multi-scale residual network for eeg-based sleep staging," 01 2021.
[47] A. Vilamala, K. Madsen, and L. Hansen, "Deep convolutional neural networks for interpretable analysis of eeg sleep stage scoring," 10 2017.
[48] R. Duggal, S. Freitas, C. Xiao, D. H. Chau, and J. Sun, "Rest: Robust and efficient neural networks for sleep monitoring in the wild," in *Proceedings of The Web Conference 2020*, 2020, pp. 1704–1714.
[49] M. Murtazina and T. Avdeenko, "An ontology-based knowledge representation in the field of cognitive functions assessment," *IOP Conference Series: Materials Science and Engineering*, vol. 919, p. 052013, 09 2020.
[50] R. Malhotra and A. Avidan, *Sleep Stages and Scoring Technique*, 12 2014, pp. 77–99.
[51] A. Hertig-Godeschalk, J. Skorucak, A. Malafeev, P. Achermann, J. Mathis, and D. R. Schreier, "Microsleep episodes in the borderland between wakefulness and sleep," *Sleep*, vol. 43, no. 1, 07 2019, zsz163. [Online]. Available: https://doi.org/10.1093/sleep/zsz163
[52] V. R. J. D. 6, "Difference between fir filter and iir filter (with comparison chart)," Feb 2021. [Online]. Available: https://circuitglobe.com/difference-between-fir-filter-and-iir-filter.html
[53] E. Sejdić, I. Djurović, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105120040800002X
[54] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
[55] Wikipedia contributors, "Welch's method — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Welch%27s_method&oldid=1070861388, 2022.
[56] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "Joint classification and prediction cnn framework for automatic sleep stage classification," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, p. 1285–1296, May 2019. [Online]. Available: http://dx.doi.org/10.1109/TBME.2018.2872652
[57] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
[58] M. X. Cohen, *Analyzing neural time series data: theory and practice*. MIT press, 2014.
[59] P. Memar and F. Faradji, "A novel multi-class eeg-based sleep stage classification system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, pp. 1–1, 11 2017.
[60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017. [Online]. Available: https://arxiv.org/abs/1706.03762
[61] X. Jiang, "Mrnet: a multi-scale residual network for eeg-based sleep staging," *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2021.
[62] M. Khalid, J. Baber, M. K. Kasi, M. Bakhtyar, V. Devi, and N. Sheikh, "Empirical evaluation of activation functions in deep convolution neural network for facial expression recognition," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, 2020, pp. 204–207.
[63] A. Malafeev, A. Hertig-Godeschalk, D. Schreier, J. Skorucak, J. Mathis, and P. Achermann, "Automatic detection of microsleep episodes with deep learning," 09 2020.

Jash Shah is a final year undergraduate student pursuing his Bachelors of Engineering from the Department of Electrical & Electronics Engineering, BITS-Pilani, Pilani Campus, 333031, India. He also worked as a visiting research intern at University of Auckland, in the Engineering Science Department. He is currently interning at NVIDIA as a Hardware Design Engineer. He has been part of various research collaborations in the past and has worked as Teaching Assistant for several courses. He shepherded the executive committee of IEEE Student Chapter at BITS Pilani. His interests lies in the intersection of Embedded Systems and EdgeML, Robotics, Computer Vision and Deep Learning, and Brain Control Interfaces.

Amit Chougule is a Ph.D. Research scholar in the Department of Electrical & Electronics Engineering, BITS-Pilani, Pilani Campus, 333031, India. He obtained his M.Tech degrees from PES University, Bangalore, India, in 2020. He is currently working as a Research scholar in the Department of Electrical & Electronics Engineering, BITS-Pilani, Pilani Campus, 333031, India. He has also worked on artificial intelligence for healthcare in MNCs such as Philips Healthcare and AIvolved Technologies. His research interests include the use of computer vision and deep learning technologies to produce artificial intelligence-based solutions for healthcare as well as autonomous driving vehicles.

Vinay Chamola is an Associate professor in the Electrical and Electronics Department, BITS) Pilani and is also a part of APPCAIR, BITS-Pilani. He received his B.E. (2010) and M.E. (2013) degrees from BITS, Pilani, and Ph.D. (2016) from the National University of Singapore. He has over 100 publications in high-ranked SCI journals, including more than 75 IEEE transactions, journal, and magazine articles. He is an Area Editor of Ad Hoc Networks, Elsevier, and IEEE Internet of Things Magazine. He also serves as Associate Editor of various journals, including IEEE Networking Letters, IEEE Consumer Electronics magazine, IET Networks, IET Quantum Communications, and so on. His research interests include the Internet of Things, 5G network provisioning, blockchain, and security.

Amir Hussain received the B.Eng. and Ph.D. degrees in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 1992 and 1997, respectively. He is currently a Professor and Director of the Centre of AI and Robotics at Edinburgh Napier University, UK. His research interests are cross-disciplinary and industry-led, aimed at developing cognitive data science and trustworthy AI technologies to engineer the smart healthcare and industrial systems of tomorrow. He has (co)authored around 600 research publications (h-index: 68), including nearly 300 journal papers and 20 Books/monographs. He has supervised over 40 Ph.D. students and led major national and international projects. He is the founding Chief Editor of Springer's Cognitive Computation journal and appointed editorial board member for Elsevier's Information Fusion and several IEEE Transactions, including on Neural Networks and Learning Systems; Artificial Intelligence; Systems, Man and Cybernetics (Systems); and Emerging Topics in Computational Intelligence. He has served as General Chair of IEEE WCCI 2020 (the world's largest technical event on computational intelligence, comprising the flagship IJCNN, FUZZ-IEEE, and IEEE CEC) and the 2023 IEEE Smart World Congress (featuring six co-located IEEE Conferences).