# A three-dimensional STFT representation for EEG alcoholism classification using 3D convolutional networks

Ayu Sekar Safitri [a,1], Achmad Rizal [a,2,*], Inung Wijayanto [a,3]

[a] School of Electrical Engineering, Telkom University, Bandung, Indonesia
[1] ayusekarass@student.telkomuniversity.ac.id; [2] achmadrizal@telkomuniversity.ac.id; [3] iwijayanto@telkomuniversity.ac.id
* corresponding author

ARTICLE INFO

**Article history**
Received December 5, 2025
Revised February 19, 2026
Accepted March 1, 2026
Available online August 31, 2026

**Keywords**
Binary classification
Convolutional neural network
Electroencephalogram
Short-time fourier transform
3D representation

ABSTRACT

Alcohol consumption can alter electrical activity in the brain, and these changes can be captured through electroencephalogram recordings. Automatically recognizing such patterns may help support early identification and reduce the reliance on subjective screening methods. In this work, a three-dimensional Short-Time Fourier Transform (3D STFT) representation is developed to encode temporal, spectral, and spatial characteristics contained in multichannel EEG data. Each electrode receives the STFT in order to produce a unique time-frequency map, which are then combined into a 3D tensor that preserves how the signal evolves across channels. Unlike 2D spectrogram, the 3D representation preserves spatial interactions across electrodes within the time-frequency domain. This multidimensional structure allows the model to interpret the EEG not as separate slices but as a unified volume. The 3D tensor is subsequently used as a 3D CNN's input for classifying EEG trials into alcoholic and non-alcoholic categories. Evaluation is performed using publicly available UCI Alcoholic EEG dataset. The proposed approach yields strong performance, producing accuracy, precision, and recall reaching 98.96%, 98.76%, and 99.17%, respectively, alongside an F1-score of 98.96 and AUC of 99.50%.. These results indicate that combining temporal, spectral, and spatial information within a single representation allows the network to extract deeper and more informative neural patterns compared to conventional 1D features or 2D spectrogram-based inputs. While STFT's fixed window length limits its ability to represent rapid or highly irregular non-stationary changes, the overall results show that the 3D representation provides a comprehensive view of EEG dynamics and enables effective classification of alcoholic versus non-alcoholic subjects.

© 2026 The Author(s).
This is an open access article under the CC–BY-SA license.

## 1. Introduction

Alcohol use continues to be a significant worldwide health concern. The World Health Organization has released figures indicating that alcohol use is responsible for around 2.6 million fatalities per year, or 4.7% of all deaths worldwide [1]. The WHO also highlights that even low levels of alcohol intake carry health risks, which escalate with higher and more prolonged consumption. These points illustrate that the consequences of alcohol misuse are not limited to physical harm but also extend to neurological and psychological domains [1]–[3].

Long-term alcohol use can cause noticeable changes in how the brain works. Alcohol interferes with key neurotransmitters that help regulate communication between brain cells, especially those involved in balancing excitation and inhibition. When this balance is repeatedly disturbed, areas of the brain that manage thinking, memory, emotional control, and decision-making start to function less effectively [2],

[4]. Research shows that regions the limbic system and the prefrontal cortex, for example can become less active or even shrink after prolonged exposure to alcohol [5]. These changes often lead to difficulties with attention, slower information processing, poorer memory, and reduced self-control. Over time, these effects participate to the development of Alcohol Use Disorder (AUD) and make it harder for individuals to manage everyday tasks and emotions [2], [6].

Changes in brain electrical activity can also demonstrate the Alcohol's effects on the anatomy and physiology of the brain [7]. From a neuroscience perspective, these electrical patterns reflect how large groups of neurons communicate in real time. Electroencephalography (EEG) provides a simple and non-invasive way to capture this activity, making it widely used to study cognitive and behavioral changes related to alcohol use [8], [9]. Studies has reported that those who suffer from AUD exhibit increased beta activity., particularly in frontal and central areas, a pattern often linked to increased cortical excitability and reduced inhibition [10]. Some findings also show elevated theta power, which can indicate slower information processing, along with reductions in alpha activity that reflect weaker inhibitory control [11]. Besides frequency patterns, alcohol-related changes are frequently seen in the frontal cortex, where both general activity and connectivity tend to drop. Since this region supports attention, decision-making, and self-regulation, reduced frontal engagement helps explain many of the cognitive and behavioral difficulties observed in long-term alcohol users [12].

EEG provides detailed information about brain activity because it captures electrical signals from many scalp locations at once, allowing to observe how neural patterns shift over time, across frequency bands, and between different brain regions [7]. A wide range of analytical approaches has been explored in studies of alcohol-related EEG abnormalities. Tlotleng et al. [13] used Power Spectral Density (PSD) to examine shifts in theta, alpha, and beta power across different levels of AUD severity. Using quantitative electroencephalography (qEEG), Bauer et al. [14] discovered that those with foetal alcohol spectrum disorders (FASD) have an elevated theta/beta ratio and predominant alpha over beta activity. Other works have emphasized spatial aspects, analyzing channel-wise differences or functional connectivity to assess how alcohol disrupts communication between brain regions. Nazri et al. [15] found reduced connectivity in AUD subjects, particularly in central and parietal areas, using Partial Directed Coherence (PDC). Similarly, Masroor et al. [16] observed decreased frontal-parietal connectivity in AUD when examining the default mode network. Long-term alcohol exposure has also been defined by time-domain characteristics including entropy, kurtosis, skewness, mean, and standard deviation, as demonstrated by Manivannan et al. [17].

Although these approaches have provided valuable insights, most existing studies rely on 1D or 2D EEG representations that examine spatial, time-frequency, frequency-domain, and time-domain aspects. either separately or in a sequential manner. Time-domain and frequency-domain features often compress dynamic neural activity into summary statistics, while conventional 2D time-frequency representations depict spectral evolution over time but do not explicitly encode spatial interactions across EEG channels. In addition, image-based EEG representations have been investigated, which show better discrimination than traditional time-domain analysis by organizing Fourier-transformed multichannel EEG data into spectral image [18]. However, EEG signals are inherently dynamic and non-stationary, with frequency content that evolves over time and varies across different scalp regions [19]. As a result, analyzing temporal, spectral, and spatial dimensions in isolation may fail to capture their interdependencies, potentially overlooking complex multichannel neural patterns associated with long-term alcohol use. These limitations highlight the need for an analytical framework capable of jointly modeling temporal evolution, spectral characteristics, and spatial distribution within a unified representation. Preserving these dimensions simultaneously allows for a more comprehensive characterization of alcohol-induced brain activity and facilitates the discovery of interaction patterns that may remain hidden when each aspect is examined independently [20]. This study addresses this gap by introducing a 3D STFT representation that jointly preserves temporal, spectral, and spatial information from multichannel EEG signals.

Time-frequency analysis is widely used to study EEG because the signal constantly changes over time and its frequency content does not stay the same. Instead of looking only at the time domain or only at the frequency domain, time-frequency methods show how different rhythms appear, disappear, or shift as the signal evolves [21]. The Short-Time Fourier Transform (STFT) is among the most often utilised methods. The spectrum of each window is computed in STFT after the EEG data is split up into brief overlapping windows. A time-frequency map that shows when particular frequency bands are stronger or weaker is created by this approach [22]. Because it is easy to use, interpret, and identify changes that occur gradually or abruptly over time, STFT is still widely used in EEG research [23].

However, the time-frequency output of STFT is usually represented in a two-dimensional form, showing only time and frequency [22]. This means the spatial information from multiple EEG channels is not directly included in the representation. To address this limitation, this study extends the conventional STFT into a three-dimensional representation by incorporating EEG channels as an additional dimension. This 3D STFT representation preserves temporal, spectral, and spatial information simultaneously, providing a richer and more informative description of alcohol-related neural activity. The suggested three-dimensional representation is then given to a Convolutional Neural Network (CNN), enabling the model to learn discriminative features across time, frequency, and channel dimensions in an integrated manner. By combining a unified EEG representation with a deep learning framework, this study aims to enhance the ability to distinguish between EEG signals from alcoholic and non-alcoholic sources while offering a more comprehensive approach for analyzing alcohol-induced brain dynamics.

Using a stacking method, this study's main contribution is the establishment of a three-dimensional STFT representation for multichannel EEG analysis. By organizing channel-wise time-frequency spectrograms into a unified 3D tensor, the proposed representation preserves temporal, spectral, and spatial channel information simultaneously without requiring handcrafted spatial features or complex connectivity measures. This representation provides a more informative input structure for subsequent learning while remaining compatible with CNN The study further demonstrates the effectiveness of this 3D stacking approach for alcoholic EEG classification using a publicly available dataset.

This paper's remainder is organized in the following manner: Section 2 outlines the techniques employed in this study, covering the dataset, preprocessing procedure, construction of the 3D STFT representation, and the CNN model used for classification. Section 3 reports the experimental findings and offers an analysis of the outcomes. The key findings are outlined in Section 4, along with possible directions for further investigation.

## 2. Method

The processing pipeline used in this investigation converts unprocessed EEG recordings of alcohol into a structured representation appropriate for deep learning categorization, as shown in Fig. 1. The STFT is used to first transform the original multichannel EEG data into two-dimensional time-frequency spectrograms. The temporal, spectral, and spatial information is then preserved by stacking these spectrograms from every channel to create a three-dimensional tensor. The resulting 3D representation is used as input to a CNN which automatically learns discriminative features related to alcoholic and non-alcoholic brain activity. Finally, the model's performance is evaluated to assess its effectiveness in distinguishing between the two groups.

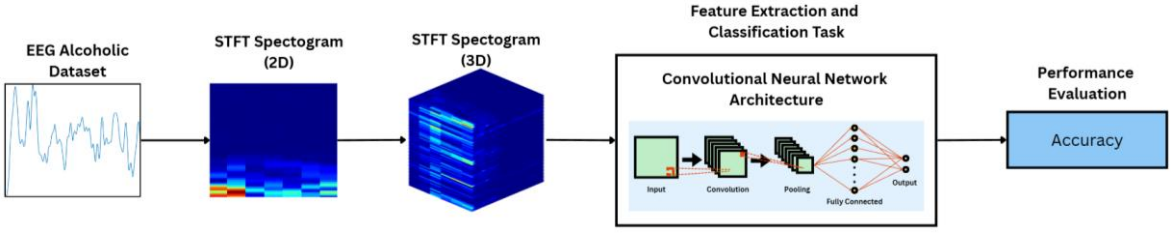



**Fig. 1.** Overview of the proposed EEG classification pipeline

### 2.1. Dataset

The EEG records were acquired from EEG Alcoholism Dataset created by H. Begleiter [24]. The dataset comes from a large-scale study that looked at the correlations between EEG and hereditary susceptibility to drinking. The recordings were collected from 64 scalp electrodes following standard electrode placement guidelines. Each signal was sampled at 256 Hz, producing 1-second EEG segments with approximately 256 samples. The dataset includes two subject groups: alcoholic and control (non-alcoholic). During the experiment, subjects were presented with visual stimuli as part of a recognition task. A total of 122 subjects participated, and every participant finished 120 trials. across different stimulus conditions. For this study, the recordings were organized into two classification categories: alcoholic and non-alcoholic, resulting in 475 alcoholic trials and 473 non-alcoholic trials. The dataset provided by UCI is already separated into training and testing subsets, consisting of 468 samples were used for testing and 480 samples for training. This predefined split is based on out-of-sample EEG trials rather than subject-wise separation. Specifically, the same subjects are present in both the training and testing sets, but different EEG runs are used for training and testing. Therefore, the evaluation reflects within-subject generalization across unseen trials rather than subject-independent classification.

### 2.2. Short-Time Fourier Transform (STFT)

STFT is frequently employed to examine signals whose frequency characteristics change over time, such as EEG recordings [22], [23]. Unlike the conventional Fourier Transform, which provides only an overall view of a signal's frequency content, STFT examines the signal in short, overlapping segments. This approach makes it possible to track when certain frequency components appear or disappear, which is essential for studying the rapidly changing patterns often found in EEG data. STFT will localize the analysis in time by applying a window function that selects a small portion of the signal. The Fourier Transform is then computed for every windowed section, and the window is slid across the entire signal to obtain a sequence of local spectra [25]. As a result, STFT produces a time-frequency representation that describes how power across different frequency bands evolves over time [25]. The STFT can be expressed as in Equation (1) [26]:

$$X(m,k) = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-\frac{j2\pi kn}{N}} \quad (1)$$

Here, $w[n-m]$ is the analysis window centered at time index $m$, $k$ signifies the frequency bin, and $N$ is the FFT length. Each coefficient $X(m,k)$ indicates how strongly a particular frequency contributes at a specific moment. An illustration of the magnitudes of these coefficients produces a spectrogram, a two-dimensional depiction of the frequency components of the signal that vary with time. A significant determinant of the compromise between resolution of frequency and time is the selected length of the window. Shorter windows allow for the detection of short-duration, fast events, but the frequency detail is reduced. On the other hand, a longer window enhances frequency discrimination but decreases the capacity to monitor rapid temporal changes [27].

### 2.3. Construction of the STFT 3D Representation

The STFT representation in this study was produced by applying the transform independently to each EEG channel. The Short-Time Fourier Transform was calculated for each channel with an FFT length of 128 and a 64-sample Hamming window with 50% overlap. At the 256 Hz sampling rate, these characteristics were chosen to provide a workable balance between time and frequency resolution. With this configuration, the STFT can record both broad oscillatory patterns that frequently show up in EEG data and quick temporal variations [28]. The 64-sample window length provides sufficient temporal resolution to capture short-term EEG dynamics, while the 50% overlap reduces spectral discontinuities and improves temporal continuity between adjacent windows. An FFT length of 128 was chosen to ensure adequate frequency resolution within the conventional EEG bands while maintaining computational efficiency [29].

The magnitude of the STFT output forms a two-dimensional time-frequency spectrogram that shows how energy in various frequency bands changes throughout the 1-second recording. This

representation highlights the local spectral characteristics of each channel. As shown in Fig. 2, lower-frequency components tend to dominate the spectrogram, and variations in color intensity indicate fluctuations in signal power over time. Because each spectrogram corresponds to a single electrode, it serves as the foundational element for constructing the final 3D representation and provides a clear visualization of the time-frequency patterns extracted from one channel.

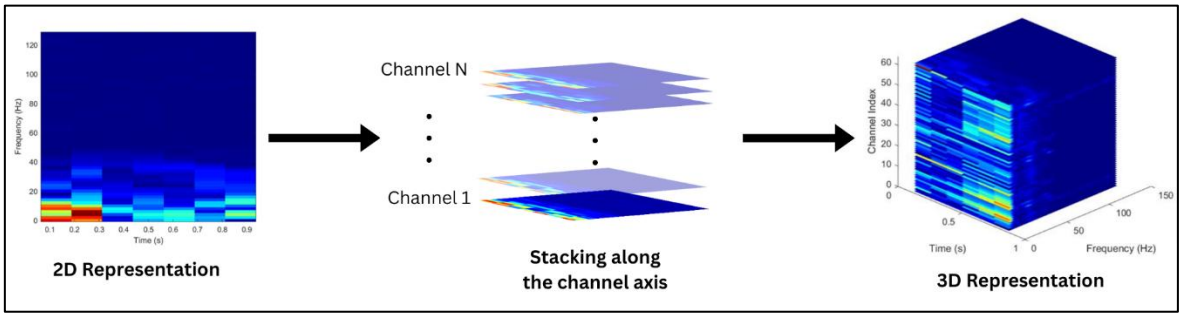



**Fig. 2.** Illustration of the 2D to 3D STFT construction process

Once a spectrogram had been generated for every channel, the individual 2D matrices were stacked along a new axis representing the electrode index. This stacking procedure produces a three-dimensional tensor with dimensions time × frequency × channel. For each 1-second EEG segment sampled at 256 Hz, the resulting 3D STFT tensor has dimensions 7 × 65 × 64, corresponding to time frames, frequency bins, and EEG channels, respectively. Each slice of the tensor captures the time-frequency structure of one electrode, while the overall cube integrates information from the entire scalp. The resulting 3D representation preserves temporal dynamics, spectral features, and spatial relationships simultaneously, allowing the CNN to learn richer and more discriminative patterns across all three dimensions. The 3D STFT tensor was formed using linear magnitude values. No log-scaling or additional normalization was applied to the STFT magnitudes prior to model input.

In general, increasing the dimensionality of a signal representation leads to higher computational and memory demands [30], [31]. In this work, the 3D STFT representation introduces manageable overhead given the short signal duration and number of EEG channels, and provides richer temporal-spectral-spatial information.

### 2.4. Convolutional Neural Network (CNN)

The 3D STFT tensors generated in the earlier step serve as the input for 3D-CNN designed to classify EEG trials into alcoholic and non-alcoholic categories. Each tensor represents the signal in three dimensions which are time, frequency, and channel, allowing the network to process it as a volumetric data block. The model begins with two convolutional stages that extract relevant features from this 3D structure. In the first stage, a 3D convolution layer with a 3 × 3 × 3 kernel and 16 filters is applied, then a 3D max-pooling layer with a 2 × 2 × 2 pooling size and stride of 2, and a RelU activation. This pattern is repeated in the second convolutional block, which increases the number of filters to 32. This enables learning in the network more intricate and complicated patterns across the time-frequency-channel dimensions. Together, these layers progressively reduce the depth of the feature maps while maintaining the most meaningful activations that distinguish alcoholic EEG characteristics [32]. The use of two convolutional stages with increasing filter depth was chosen to progressively capture low-level and higher-level patterns across the time-frequency-channel dimensions while avoiding excessive model complexity and overfitting, given the dataset size [33].

Following the convolutional procedures, the generated feature maps are transferred to a fully connected layer of 128 neurones after being flattened. In order to reduce overfitting during training and enhance generalisation, a ReLU activation and a dropout rate of 0.5 are employed. A softmax layer processes the output units for the alcoholic and non-alcoholic classes produced by the last dense layer to provide likelihood scores. In training, the loss is calculated using a classification layer [34], [35]. The entire model design is illustrated in Fig. 3, with a detailed summary of the 3D-CNN architecture, including layer types, kernel sizes, and output dimensions, is provided in Table 1. The model's

classification capacity was thoroughly assessed using a set of standard performance criteria that are frequently employed in classification tasks.

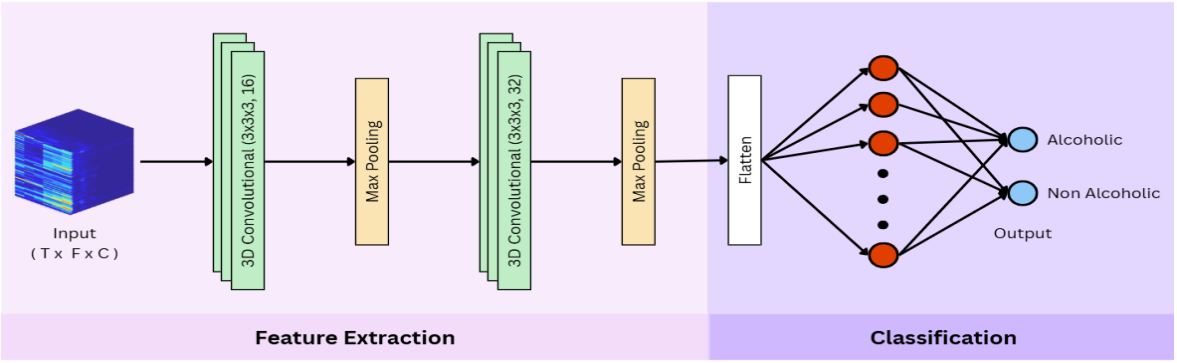



**Fig. 3.** Proposed 3D-CNN Architecture

**Table 1.** Summary of the proposed 3D-CNN architecture

| Layer | Kernel/Units | Stride | Output Size |
|---|---|---|---|
| Input | - | - | 7 × 65 × 64 |
| Conv3D | 3 × 3 × 3, 16 | 1 | 7 × 65 × 64 × 16 |
| MaxPool3D | 2 × 2 × 2 | 2 | 3 × 32 × 32 × 16 |
| Conv3D | 3 × 3 × 3, 32 | 1 | 3 × 32 × 32 × 32 |
| MaxPool3D | 2 × 2 × 2 | 2 | 1 × 16 × 16 × 32 |
| Fully Connected | 128 | - | 128 |
| Dropout | 0.5 | - | 128 |
| Fully Connected | 2 | - | 2 |

### 2.5. Performance Evaluation

The classifier's performance was assessed according to a number of criteria, such as accuracy, precision, recall, and the F1-score. Equation (2) illustrates accuracy, which is the frequency with which the model generates accurate predictions for every test sample [36]. Equation (3) defines precision as the model's ability to detect alcoholic EEG trials without producing a high number of false positives [37]. Equation (4)'s recall shows how well the model can identify alcoholic samples that need to be categorised as positive [37]. When both accuracy and recall are significant, the F1-score in Equation (5) is useful since it aggregates the two into a single number [36].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1Score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{5}$$

These evaluation metrics rely on four fundamental classification outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The model's response when deciding whether an EEG trial falls into the alcoholic or non-alcoholic class is demonstrated by these values taken together. Also used to supplement these numerical measurements is a confusion matrix. It gives a clear visual overview of the model's predictions by presenting the proportion of samples from each class that were properly or mistakenly recognised. This illustration provides more information about the model's performance traits and aids in identifying any recurrent misclassification trends [38].

In addition to the commonly used performance metrics, this study also provides the classifier's Area Under the ROC Curve (AUC) for evaluation without relying on a fixed decision threshold. As the classification threshold is changed, the ROC curve shows how the true positive rate (TPR) changes in proportion to the false positive rate (FPR). The integral of the ROC curve across the interval [0, 1] is used to get the AUC value, as expressed in Equation (6) [39]. This measure indicates how effectively the model can distinguish between the two classes regardless of the threshold applied, offering a broader understanding of its overall discriminative strength [39], [40].

$$AUC = \int_0^1 ROC(u)du = \int_0^1 TPR(FPR)d(FPR) \quad (6)$$

The experiments were conducted using a single fixed training-testing split provided by the dataset, and cross-validation was not applied. Consequently, statistical variability across folds was not evaluated; however, all experiments were performed using the same data split, network architecture, training settings, and evaluation protocol to ensure consistent evaluation conditions.

## 3. Results and Discussion

The performance evaluation focuses on assessing how well the proposed 3D STFT representation, combined with a 3D CNN architecture, can distinguish between the two classes in the dataset. The overall accuracy of the model reached 98.96%, indicating that nearly all samples in the test set were classified correctly. This high value suggests that the combination of 3D STFT features and 3D convolutional layers was able to extract meaningful patterns from the EEG data. Because EEG signals are often noisy and contain overlapping information across channels and frequencies [41], [42], achieving a high accuracy shows that the model captured key structures in the time-frequency-space representation. It also implies that the preprocessing steps, such as computing STFT per channel and stacking them into a 3D tensor, were effective for representing the signal in a way that the CNN could interpret clearly.

A closer examination of the confusion matrix in Fig. 4 provides additional insight into how the model handled each class. For the alcoholic class, the model correctly identified 238 out of 240 samples, misclassifying only 2. For the control group, the model correctly predicted 237 samples and made 3 errors. These small misclassification counts indicate consistent performance across both classes, with no strong class imbalance or directional bias. The strong diagonal dominance in the confusion matrix reflects that the learned features align well with the differences required to discriminate the two EEG categories.

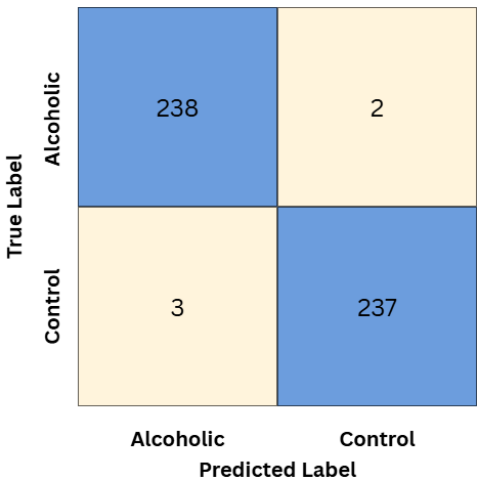



**Fig. 4.** Confusion Matrix Result

The model also achieved a precision score of 98.76%, demonstrating its ability to minimize false positives. When the classifier assigns a sample to the alcoholic category, this prediction is correct in

almost every case. At the same time, the model correctly detected almost all of the alcoholic samples in the dataset, as seen by the recall value of 99.17%. The classifier consistently performs well across all dimensions, as seen by its F1-score of 98.96, which reflects the harmonic balance between accuracy and recall. A high F1-score indicates the capacity of the model to manage false negatives and false positives. This reinforces the effectiveness of combining 3D feature encoding with 3D convolutional learning, enabling the network to utilize the full richness of the multi-dimensional EEG structure. The ROC curves shown in Fig. 5 offer another perspective on classifier performance. In this plot, Class 1 corresponds to the alcoholic group, while Class 2 represents the control group. Both classes produce curves that rise sharply toward the upper-left corner, with an AUC of 99.50%. An AUC value close to 1.0 indicates excellent discriminative power across a wide range of probability thresholds. This suggests that even if the decision boundary were adjusted, the model would still be able to separate alcoholic and non-alcoholic EEG signals with very few ambiguous or overlapping cases.

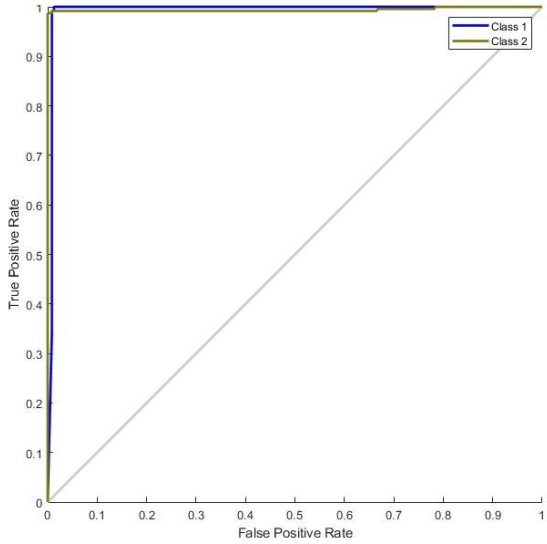



**Fig. 5.** ROC curve for proposed model

To further evaluate the benefit of the proposed 3D representation, baseline experiments were conducted under the same fixed train-test split using 1D and 2D CNN models as presented in Table 2. The 1D time-domain CNN achieved an accuracy of 64.79%, with precision and recall values of 64.79% and an AUC of 0.7491, indicating limited discriminative capability when multichannel EEG signals are processed only in the temporal domain. This result suggests that compressing EEG recordings into a purely one-dimensional representation leads to substantial information loss, particularly in terms of spectral and spatial characteristics. When a 2D STFT representation was employed, performance improved to 79.79% accuracy, with precision and recall of 79.79% and an AUC of 0.9187, demonstrating the clear benefit of incorporating time-frequency information.

In contrast, the proposed 3D STFT combined with a 3D CNN achieved substantially higher performance, reaching 98.96% accuracy, 98.76% precision, 99.17% recall, and an AUC of 0.9950. These results indicate not only improved overall accuracy but also a more balanced handling of false positives and negatives. The strong recall value shows that the model successfully identifies nearly all alcoholic EEG samples, while the high precision confirms reliable discrimination from control signals. This improvement highlights the importance of explicitly preserving spatial information across EEG channels, enabling the network to capture inter-channel relationships and coordinated neural activity patterns that are not accessible in 1D or 2D representations.

The strong performance of the proposed method can be partly explained by the inclusion of spatial information across EEG channels in the 3D STFT representation. By stacking time-frequency spectrograms from all channels, the model can capture relationships between different electrodes that are not explicitly represented in 1D or 2D features. This spatial arrangement allows the network to

observe how activity in different brain regions evolves together over time and frequency. Such joint information is particularly important for EEG signals, where meaningful patterns often arise from interactions across multiple channels rather than isolated electrodes. Consequently, the preservation of spatial context in the 3D STFT representation contributes to more robust and reliable classification performance.

**Table 2.** Performance comparison of the suggested approach with baseline models

| Representation | Classifier | Accuracy (%) | Precision (%) | Recall (%) | F1-Score | AUC |
|---|---|---|---|---|---|---|
| 1D time-domain | 1D CNN | 64.79 | 64.79 | 64.79 | 0.6396 | 0.7491 |
| 2D STFT | 2D CNN | 79.79 | 79.79 | 79.79 | 0.7948 | 0.9187 |
| **3D STFT (Proposed)** | **3D CNN** | **98.96** | **98.76** | **99.17** | **0.9896** | **0.9950** |

A comparison with previous research, as presented in Table 3, demonstrates that a wide range of approaches have achieved strong performance in detecting alcoholic EEG patterns, each emphasizing different aspects of the signal. The suggested approach demonstrates competitive performance, achieving 98.96% accuracy, 98.76% precision, 99.17% recall, a 98.96 F1-score, and a 99.50% AUC. Although the compared studies are based on the same EEG Alcoholism dataset, they differ in preprocessing strategies, feature representations, and classification models, which can influence the reported performance. Unlike most previous studies that rely on 1D statistical descriptors or 2D time-frequency representations, the 3D STFT approach preserves temporal, spectral, and spatial channel information simultaneously. Because of its richer representation, the 3D CNN can learn multi-dimensional EEG patterns that previous feature extraction algorithms might not have been able to completely capture. As seen by the consistently high metrics found in this work, adding spatial channel structure to the time-frequency domain offers an alternative viewpoint to current methods and may help provide more reliable classification.

**Table 3.** Comparison with previous studies

| Reference | Methods | Result |
|---|---|---|
| Manekar et al. [45] | DWT + statistical features | Accuracy: 95.6%, Sensitivity: 91.3%, F1: 95.5% |
| Purnamasari et al. [46] | Relative Wavelet Bispectrum + Bispectrum-Gaussian | Accuracy: 90% |
| Shen et al. [47] | CWT + Mutual Information connectivity | Accuracy: 96.25%, F1: 98.06% |
| Mohd Nazri et al. [15] | Partial Directed Coherence | Accuracy: 96.37% |
| Sadiq et al. | Fractional Fourier Transform (FrFT) | Accuracy: 97.5%, Sensitivity: 96.7%, Specificity: 98.3% |
| **Proposed Method** | **3D STFT** | **Accuracy: 98.96%, Precision: 98.76%, Recall: 99.17%, F1-Score: 98.96, AUC: 99.50%** |

Although the suggested approach achieves an accuracy of 98.96%, it is important to note that the use of STFT comes with inherent limitations. Because STFT employs a fixed window size throughout the entire signal, it is less flexible for analyzing non-stationary EEG patterns whose temporal characteristics vary in duration [43]. This fixed-window design introduces a resolution trade-off, where improving temporal resolution leads to reduced frequency resolution and vice versa. As a result, rapid transient events or subtle oscillatory changes may not be captured optimally, potentially limiting the richness of the time-frequency representation [43], [44]. Despite these constraints, the constructed 3D STFT tensor still preserves the key temporal, spectral, and spatial information needed for effective

classification. In future work, adaptive or multi-resolution time-frequency methods, such as wavelet-based transforms, could address this limitation by adjusting time-frequency resolution across different frequency bands, allowing both short-duration transients and longer oscillatory components to be represented more effectively.

## 4. Conclusion

This study introduced a three-dimensional STFT representation that integrates temporal, spectral, and spatial information from multichannel EEG recordings, and demonstrated its effectiveness for classifying alcoholic and non-alcoholic brain signals using a 3D-CNN framework. Main contribution of this work lies in explicitly preserving channel-wise spatial information within the time-frequency representation, enabling joint learning across time, frequency, and EEG channels. With an F1-score of 98.96, an accuracy of 98.96%, precision of 98.76%, recall of 99.17%, and an AUC of 99.50%, the suggested approach produced excellent results. These findings demonstrate that deep learning combined with a richer 3D representation allows the model to identify brain patterns that conventional 1D or 2D feature extraction methods would miss. The results also suggest that keeping the spatial arrangement of EEG channels helps improve classification because the model can observe how patterns appear across different parts of the scalp. Treating the STFT output as a 3D input makes the learned features richer compared to using only 1D or 2D representations. Although the 3D STFT representation demonstrates strong classification capability, the fixed window size in STFT presents certain limitations for analyzing highly non-stationary EEG dynamics. Adaptive or multi-resolution methods, such as wavelet-based approaches, may help capture transient or rapidly evolving patterns more effectively. Future research may also explore larger datasets, subject-independent evaluation, and the incorporation of explainable artificial intelligence (XAI) methods to provide deeper insight into the neural features that drive the model's decisions.

### Acknowledgment

Telkom University is acknowledged by the authors for its assistance and facilities provided through the Signal Processing for Intelligent System Laboratory. The financial assistance, guidance, and administrative coordination offered throughout the research process played an important role in enabling the accomplishment of the task of this study.

### Declarations

**Author contribution.** Ayu Sekar Safitri conducted the simulations, performed the data analysis, and write the manuscript. Achmad Rizal provided the dataset, contributed to the conceptual direction of the study, and offered critical feedback throughout the writing process. Inung Wijayanto helped with the manuscript's editing and improvement in addition to helping with data analysis. The final draft of the work has been reviewed and approved by all authors, who also consent to its publication
**Funding statement.** There was no outside support for this study
**Conflict of interest.** No conflicts of interest are disclosed by the writers.
**Additional information.** There is no further information available for this paper.

### Data and Software Availability Statements

This study uses secondary EEG data derived from the publicly available Alcoholic EEG Dataset provided through the UCI Machine Learning Repository. The dataset can be accessed through its official DOI link: https://doi.org/10.24432/C5TS3D. No restrictions apply, and no additional data were generated by the authors.

### References

[1] "Global status report on alcohol and health and treatment of substance use disorders," *World Health Organization*, 2024. [Online]. Available at: https://www.who.int/publications/i/item/9789240096745.

[2] W. Yang, R. Singla, O. Maheshwari, C. J. Fontaine, and J. Gil-Mohapel, "Alcohol Use Disorder: Neurobiology and Therapeutics," *Biomedicines*, vol. 10, no. 5, p. 1192, May 2022, doi: 10.3390/biomedicines10051192.

[3] K. Renu *et al.*, "Molecular mechanisms of alcohol's effects on the human body: A review and update," *J. Biochem. Mol. Toxicol.*, vol. 37, no. 12, p. e23502, Dec. 2023, doi: 10.1002/jbt.23502.

[4] M. Matuszewska *et al.*, "The Long-Term Effects of Alcohol on the Central Nervous System: Mechanisms, Cognitive Decline, and Associated Disorders," *J. Educ. Heal. Sport*, vol. 80, p. 59611, Apr. 2025, doi: 10.12775/JEHS.2025.80.59611.

[5] S. Maharjan *et al.*, "Executive Dysfunction in Patients With Alcohol Use Disorder: A Systematic Review," *Cureus*, p. 14, Sep. 2022, doi: 10.7759/cureus.29207.

[6] M. Roberge, M. Boisvert, and S. Potvin, "A Meta-Analysis of Task-Based fMRI Studies on Alcohol Use Disorder," *Brain Sci.*, vol. 15, no. 7, p. 665, Jun. 2025, doi: 10.3390/brainsci15070665.

[7] C. E. Fairbairn, D. Kang, and K. D. Federmeier, "Alcohol and Neural Dynamics: A Meta-analysis of Acute Alcohol Effects on Event-Related Brain Potentials," *Biol. Psychiatry*, vol. 89, no. 10, pp. 990–1000, May 2021, doi: 10.1016/j.biopsych.2020.11.024.

[8] H. Zhang, J. Yao, C. Xu, and C. Wang, "Targeting electroencephalography for alcohol dependence: A narrative review," *CNS Neurosci. Ther.*, vol. 29, no. 5, pp. 1205–1212, May 2023, doi: 10.1111/cns.14138.

[9] T. S. Bel-Bahar, A. A. Khan, R. B. Shaik, and M. A. Parvaz, "A scoping review of electroencephalographic (EEG) markers for tracking neurophysiological changes and predicting outcomes in substance use disorder treatment," *Front. Hum. Neurosci.*, vol. 16, pp. 1–31, Oct. 2022, doi: 10.3389/fnhum.2022.995534.

[10] R. L. Hack, M. Aigner, M. Musalek, R. Crevenna, and L. Konicar, "Brain regulation training improves emotional competences in patients with alcohol use disorder," *Soc. Cogn. Affect. Neurosci.*, vol. 19, no. 1, pp. 1–14, Aug. 2024, doi: 10.1093/scan/nsae048.

[11] E. López-Caneda, F. Cadaveira, A. Correas, A. Crego, F. Maestú, and S. Rodríguez Holguín, "The Brain of Binge Drinkers at Rest: Alterations in Theta and Beta Oscillations in First-Year College Students with a Binge Drinking Pattern," *Front. Behav. Neurosci.*, vol. 11, pp. 1–10, Sep. 2017, doi: 10.3389/fnbeh.2017.00168.

[12] Y. Liu *et al.*, "Resting-state EEG, Substance use and Abstinence After Chronic use: A Systematic Review," *Clin. EEG Neurosci.*, vol. 53, no. 4, pp. 344–366, Jul. 2022, doi: 10.1177/15500594221076347.

[13] K. M. Tlotleng and R. S. Jamisola, "An Analysis of the Severity of Alcohol Use Disorder Based on Electroencephalography Using Unsupervised Machine Learning," *Big Data Cogn. Comput.*, vol. 9, no. 7, p. 170, Jun. 2025, doi: 10.3390/bdcc9070170.

[14] W. Bauer *et al.*, "Initial study on quantitative electroencephalographic analysis of bioelectrical activity of the brain of children with fetal alcohol spectrum disorders (FASD) without epilepsy," *Sci. Rep.*, vol. 13, no. 1, p. 109, Jan. 2023, doi: 10.1038/s41598-022-26590-4.

[15] A. K. Mohd Nazri *et al.*, "Partial directed coherence analysis of resting-state EEG signals for alcohol use disorder detection using machine learning," *Front. Neurosci.*, vol. 18, pp. 1–15, Jan. 2025, doi: 10.3389/fnins.2024.1524513.

[16] K. Masroor, D. M. Khan, H. R. Khan, N. Yahya, K. Arshad, and K. Assaleh, "Deciphering Default Mode Network Dynamics in Alcohol Use Disorder through Directed Transfer Function Analysis of EEG," *IEEE Access*, pp. 1–1, 2024, doi: 10.1109/ACCESS.2024.3518916.

[17] L. J. Aradhana Manekar, "Automated Classification of Alcoholism Using Discrete Cosine Harmonic Wavelet-Packet Transform," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 10s, pp. 125–136, Feb. 2025, doi: 10.52783/jisem.v10i10s.1361.

[18] A. Rizal, D. S. Beu, and S. Ziani, "Alcoholic EEG Classification Multichannel FFT Image," 2025, pp. 473–478, doi: 10.1007/978-3-031-90921-4_66.

[19] P. Croce, A. Quercia, S. Costa, and F. Zappasodi, "EEG microstates associated with intra- and inter-subject alpha variability," *Sci. Rep.*, vol. 10, no. 1, p. 2469, Feb. 2020, doi: 10.1038/s41598-020-58787-w.

[20] A. Raj, P. Verma, and S. Nagarajan, "Structure-function models of temporal, spatial, and spectral characteristics of non-invasive whole brain functional imaging," *Front. Neurosci.*, vol. 16, pp. 1–11, Aug. 2022, doi: 10.3389/fnins.2022.959557.

[21] Z. Zhang, "Spectral and Time-Frequency Analysis," in *EEG Signal Processing and Feature Extraction*, Singapore: Springer Singapore, 2019, pp. 89–116, doi: 10.1007/978-981-13-9113-2_6.

[22] G. Lekkas, E. Vrochidou, and G. A. Papakostas, "Time–Frequency Transformations for Enhanced Biomedical Signal Classification with Convolutional Neural Networks," *BioMedInformatics*, vol. 5, no. 1, p. 7, Jan. 2025, doi: 10.3390/biomedinformatics5010007.

[23] A. F. Almurshedi, "Envelope analysis and short time Fourier transform of EEG for epilepsy diagnosis enhancement," *Rom. J. Neurol.*, vol. 24, no. 1, pp. 35–41, Mar. 2025, doi: 10.37897/RJN.2025.1.12.

[24] "EEG Database," *UCI Machine Learning Repository*, 1999. [Online]. Available at: https://archive.ics.uci.edu/dataset/121/eeg+database.

[25] C. Liu, P. Han, and J. Kang, "Integrating short-time linear canonical transform and joint space-time-frequency analysis for advanced representation of subsurface information in ground penetrating radar," *Sci. Rep.*, vol. 15, no. 1, p. 23315, Jul. 2025, doi: 10.1038/s41598-025-06795-z.

[26] P. Ewert, B. Wicher, and T. Pajchrowski, "Application of the STFT for Detection of the Rotor Unbalance of a Servo-Drive System with an Elastic Interconnection," *Electronics*, vol. 13, no. 2, p. 441, Jan. 2024, doi: 10.3390/electronics13020441.

[27] M. Zhang, "Multi-resolution short-time Fourier transform providing deep features for 3D CNN to classify rolling bearing fault vibration signals," *Eng. Res. Express*, vol. 6, no. 3, p. 035201, Sep. 2024, doi: 10.1088/2631-8695/ad5d52.

[28] H. Jeon, Y. Jung, S. Lee, and Y. Jung, "Area-Efficient Short-Time Fourier Transform Processor for Time–Frequency Analysis of Non-Stationary Signals," *Appl. Sci.*, vol. 10, no. 20, p. 7208, Oct. 2020, doi: 10.3390/app10207208.

[29] J. Chmiel, J. Nadobnik, S. Smerdel, and M. Niedzielska, "Neural Correlates of Huntington's Disease Based on Electroencephalography (EEG): A Mechanistic Review and Discussion of Excitation and Inhibition (E/I) Imbalance," *J. Clin. Med.*, vol. 14, no. 14, p. 5010, Jul. 2025, doi: 10.3390/jcm14145010.

[30] Z. Wang, L. Chen, P. Xiao, L. Xu, and Z. Li, "Enhancing time-frequency resolution via deep-learning framework," *IET Signal Process.*, vol. April, pp. 1–19, 2023, doi: 10.1049/sil2.12210.

[31] S. Kumawat, M. Verma, Y. Nakashima, and S. Raman, "Depthwise Spatio-Temporal STFT Convolutional Neural Networks for Human Action Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4839–4851, Jul. 2020, doi: 10.1109/TPAMI.2021.3076522.

[32] I. Rakhmatulin, M.-S. Dao, A. Nassibi, and D. Mandic, "Exploring Convolutional Neural Network Architectures for EEG Feature Extraction," *Sensors*, vol. 24, no. 3, p. 877, Jan. 2024, doi: 10.3390/s24030877.

[33] C. Jannu and S. D. Vanambathina, "Multi-stage Progressive Learning-Based Speech Enhancement Using Time–Frequency Attentive Squeezed Temporal Convolutional Networks," *Circuits, Syst. Signal Process.*, vol. 42, no. 12, pp. 7467–7493, Dec. 2023, doi: 10.1007/s00034-023-02455-7.

[34] B. Wang *et al.*, "Automatic epileptic seizure detection based on EEG using a moth-flame optimization of one-dimensional convolutional neural networks," *Front. Neurosci.*, vol. 17, pp. 1–16, Dec. 2023, doi: 10.3389/fnins.2023.1291608.

[35] S. Siuly, Y. Guo, O. F. Alcin, Y. Li, P. Wen, and H. Wang, "Exploring deep residual network based features for automatic schizophrenia detection from EEG," *Phys. Eng. Sci. Med.*, vol. 46, no. 2, pp. 561–574, Jun. 2023, doi: 10.1007/s13246-023-01225-8.

[36] G. D. Paladugu, H. K. Goru, J. M. Gundu, P. Medisetty, and K. K. Kumari, "Alcoholism Identification From Variational Mode Decomposed EEG Signal Using Machine Learning Techniques," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–7, doi: 10.1109/ICCCNT56998.2023.10307977.

[37] A. Ebrahimi, U. K. Wiil, A. Naemi, M. Mansourvar, K. Andersen, and A. S. Nielsen, "Identification of clinical factors related to prediction of alcohol use disorder from electronic health records using feature selection methods," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, p. 304, Nov. 2022, doi: 10.1186/s12911-022-02051-w.

[38] W. Mumtaz, P. L. Vuong, L. Xia, A. S. Malik, and R. B. A. Rashid, "An EEG-based machine learning method to screen alcohol use disorder," *Cogn. Neurodyn.*, vol. 11, no. 2, pp. 161–171, Apr. 2017, doi: 10.1007/s11571-016-9416-y.

[39] T. Yang and Y. Ying, "AUC Maximization in the Era of Big Data and AI: A Survey," *ACM Comput. Surv.*, vol. 55, no. 8, pp. 1–37, Aug. 2023, doi: 10.1145/3554729.

[40] J. Y. Verbakel *et al.*, "ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models," *J. Clin. Epidemiol.*, vol. 126, pp. 207–216, Oct. 2020, doi: 10.1016/j.jclinepi.2020.01.028.

[41] M. Gabardi, A. Saibene, F. Gasparini, D. Rizzo, and F. A. Stella, "A multi-artifact EEG denoising by frequency-based deep learning," in *ISSN 1613-0073 CEUR Workshop Proceedings*, 2023, pp. 1–14, [Online]. Available at: https://ceur-ws.org/Vol-3576/paper4.pdf.

[42] W. Yan, Q. Luo, and C. Du, "Channel component correlation analysis for multi-channel EEG feature component extraction," *Front. Neurosci.*, vol. 19, pp. 1–11, Apr. 2025, doi: 10.3389/fnins.2025.1522964.

[43] Y. Yan, S. B. Samdin, and K. N. Minhad, "Window-based Time-Frequency Methods for Analyzing Epileptic EEG Signals," in *2022 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Dec. 2022, pp. 292–297, doi: 10.1109/IECBES54088.2022.10079259.

[44] W. Yan and Y. Wu, "A time-frequency denoising method for single-channel event-related EEG," *Front. Neurosci.*, vol. 16, pp. 1–14, Nov. 2022, doi: 10.3389/fnins.2022.991136.

[45] A. Manekar and L. Jolly, "Wavelet Decomposition based Automated Alcoholism Classification using EEG Signal," in *2022 IEEE Bombay Section Signature Conference (IBSSC)*, Dec. 2022, pp. 1–6, doi: 10.1109/IBSSC56953.2022.10037362.

[46] P. D. Purnamasari, F. H. Zulkarnaen, M. Melinda, E. P. Sinulingga, F. Fahmi, and A. A. P. Ratna, "Alcoholic EEG Signal Feature Extraction Based on Relative Wavelet Bispectrum and Bispectrum-Gaussian Using CNN and ANN Classifier," in *2024 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Dec. 2024, pp. 1–6, doi: 10.1109/IECBES61011.2024.10990911.

[47] M. Shen, P. Wen, B. Song, and Y. Li, "Detection of alcoholic EEG signals based on whole brain connectivity and convolution neural networks," *Biomed. Signal Process. Control*, vol. 79, p. 104242, Jan. 2023, doi: 10.1016/j.bspc.2022.104242.