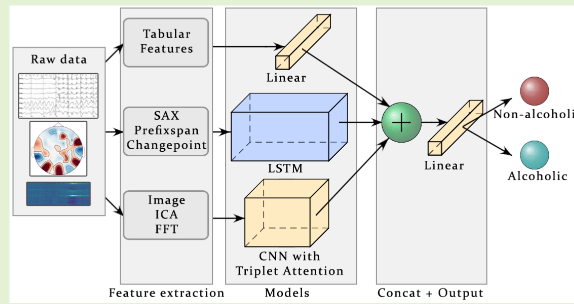
Ensemble Learning For Alcoholism Classification Using EEG Signals

Seffi Cohen, Or Katz, Dan Presil, Ofir Arbili, and Lior Rokach

***Abstract*—Excessive drinking is a major risk factor that leads to many health complications. The diagnosis of alcoholism is challenging, especially when the standard diagnostic tests rely on blood tests and questionnaires that are subjective to the patient and the examiner. The study's major goal is to find new EEG classification methods to improve past findings and construct a robust EEG classification algorithm to generate accurate predictions with explainable results. The EEG records were examined from two different perspectives and combined with an ensemble of classification models. The first approach was temporal data, and the second was images derived from the original signals. Using fast Fourier transform (FFT) and independent component analysis (ICA), we convert 64-channel temporal data into images along with applying the Symbolic Aggregate approXimation (SAX) technique. Our model combines input data in tabular, temporal, and image formats with an ensemble of linear neural networks, long short-term memory (LSTM), and efficient-net classification models. We have evaluated our method using a publicly available dataset for EEG classification of alcoholic and non-alcoholic subjects. Overall, our algorithm's highest cross-validation classification accuracy is 85.5% compared to the state-of-the-art EEG-NET's accuracy of 81.19%.**



***Index Terms*—EEG, Alcoholism Classification, Ensemble Methods**

I. Introduction

ALCOHOL consumption, especially heavy drinking, is a major risk factor for many health problems and an important factor in the global disease burden. Alcohol is the leading cause of over 30 conditions and helps many others. The most common types of alcohol-related illnesses, whether complete or partial, include cancer, diabetes, pregnancy complications, neurological disorders, heart disease, liver, and pancreatic disease, and accidental and deliberate injuries [1]. There is no standardized and sufficiently accurate method to detect alcohol abuse disorder in humans. Specialized medical personnel can physically examine symptoms and subsequent diagnostic tests to give reliable results. Unfortunately, these procedures are long and, for the most part, are still affected by human mistakes [2]. Electroencephalography (EEG) is a noninvasive technology for recording the macroscopic activity of the brain's surface layer based on an electrogram recorded from the scalp. EEG is widely used to measure the brain's electrical activity by connecting electrodes to small metal plates and wires. Visual inspection of EEG records was used in early attempts to analyze EEG data. The amount of data obtained from a single-patient study has expanded rapidly since the introduction of EEG recordings. As a result, pattern classification techniques-based automation has been used with great success [3]. The brain neurons generate small electrical charges. When multiple neurons are activated, their electrical activity can be observed and recorded with small electrodes placed on the scalp. When we record these signals, we can illustrate them with brain waves, and each electrode produces a different brain wave pattern [4]. Medical practitioners use EEG signals to detect Alcohol Abuse Disorder (AUD) in subjects. Alcoholics and non-alcoholics have been known to produce significant differences in patterns of brain activity [4]. Manually analyzing EEG signals is a challenging task. Only highly trained physicians well versed in the subject can interpret these signals and find helpful examples to help them draw specific conclusions about whether a subject has an AUD [2]. This type of EEG is remarkably noise sensitive. Data with a high noise rate make the manual classification process time-consuming and challenging, even for highly trained practitioners. As a result, there is a need for automated methods for analyzing and classifying these EEG signals that are faster, less prone to human mistakes, and more accurate [4].

In recent years, several research studies have been conducted on identifying alcoholism using EEG signals. Machine learning algorithms and statistical techniques are extensively studied for AUD classification using EEG signal, in particular, [5]–[8], [8]–[11]. Most of the previous studies utilizing machine learning techniques that manually generated features

All authors are with the Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Ben-Gurion Blvd 1, Be'er-Shea, Israel
S. C. e-mail: seffi@post.bgu.ac.il
O. K. e-mail: ork@post.bgu.ac.il
D. P. e-mail: danpr@post.bgu.ac.il
O. A. e-mail: arbili@post.bgu.ac.il
L. R. e-mail: liorrk@post.bgu.ac.il

require domain knowledge and are chosen by researchers based on their expertise. The patterns observed with EEG when a subject is resting are random. In tests done on multiple subjects, where we would want to compare their brain waves, all subjects are asked to perform a similar simple action (like staring at an image) to activate specific brain regions. The brain wave is usually divided into frequency bands: delta (1–3 Hz), theta (3.5–7.5 Hz), alpha (8.0–11.5 Hz), beta (12–28 Hz), and gamma (28.5–50.0 Hz), and the different bands have different levels of brain activity. The alpha waves are prominent in normal awake persons in a quiet and pleasant environment, beta waves occur during intense mental activity, theta waves in children and during emotional stress in adults, and delta waves in deep sleep. The theta band shows an increase in power in all the electrodes for a resting state for the alcoholic group, the alpha band has an irregular rhythm for the alcoholic group, and the beta band shows an increase in power for alcoholics compared to the control group [4]. Researchers initially focused on filtering the EEG signals, extracting features through signal processing techniques such as wavelets [12] and Fourier transforms [13], and then using traditional machine learning models to classify the signals. Bajaj *et al.* [5] obtained time-frequency images using spectrograms of the Short-Time Fourier Transform (STFT) and then classified them with Non-Negative Least Squares (NNLS). According to Zhu *et al.* [14], horizontal visibility graph entropy (HVGE) was used as an extractor of features, utilizing the KNN algorithm and the Support Vector Machine (SVM) classifier. Kumar *et al.* [7] applied fast Fourier transformation (FFT) and separated the signal into five different bands at the pre-processing stage; then, three feature extraction methods were used: absolute power, relative power, and peak power frequency. The granularity of these features was reduced and normalized by Linear Discrimination Analysis (LDA) and then classified by support vector machine (SVM) and fuzzy C-mean clustering. Acharya *et al.* [7] used SVM to classify EEG signals based on various features, including Approximate Entropy, Sample Entropy, and Largest Lyapunov Exponent. Rachman *et al.* [8] used ICA first as a noise reduction method, and then a stationary wavelet transform was used to extract features and avoid downsampling. Four layers of Probabilistic Neural Network (PNN) are used as a third step as a supervised training method. Fust *et al.* [9] used energy measures and Wavelet Packet Decomposition (WPD) for feature extraction to classify EEG signals. Ekaputri *et al.* [10] also used Wavelet, who selected MWPE for its previous promising performance in lung sound analysis, EEG analysis, and speech analysis. Using Wavelet Packet Decomposition, the EEG signal was decomposed into 2N subbands with equal bandwidth, and then the entropy was calculated as a feature, and support Vector Machine was used to classify the signal. The highest accuracy yield, 77.8%, was achieved with quadratic SVM with decomposition level 6, a low result compared to other studies. The weakness of this study was that the EEG signals from 64 channels were combined. Also, only a subset of the signal was taken for the feature extraction. E. Malar *et al.* [11] decompose the EEG data with Wavelet Transform (WT) into sub-frequency bands. Since EEG waves are non-stationary, they use wavelet decomposition to give accurate frequency information at low frequencies and accurate time information at high frequencies. The low-frequency range is considered the useful section of the EEG spectrum, and the higher frequencies are considered noise; then, the wavelet features of the EEG signals were selected and used with the Extreme Learning Machine (ELM) classifier. A different approach for feature extraction by Zubair *et al.* [6] aimed at developing new features based on the Sliding-SSA algorithm. The authors used the extracted features in various machine-learning models, including (SVM, KNN, ANN, AdaBoost, and XGBoost). All the abovementioned methods include classical machine learning techniques and manually generated features, which researchers choose based on their expertise. The following studies, which rely on deep neural networks, have generally achieved higher accuracy results. Deep neural networks (DNN) for alcoholic classification were first utilized by Farsi *et al.* [15]. The main innovation in their paper was the use of deep learning-based techniques when previous papers showed lower results using classical machine learning-based techniques. They present two deep learning-based methods for classification. The first method is based on Principal Component Analysis (PCA) for a hand-crafted feature extraction method and Artificial Neural Network (ANN) as a deep learning method. In this method, the PCA is used to extract important feature values from EEG data which are used as input to the ANN method. This method achieves an average classification accuracy of 86%. The second method is based on the Long Short-Term Memory (LSTM) network, where the raw EEG data is directly used as input to an LSTM network. The proposed LSTM-based algorithm yields better classification performances (93.00% on random train-test split) in terms of accuracy criteria compared to the reported existing methods in the literature. This approach was expanded by Fayyaz *et al.* [16], who used LSTM to classify EEG signals after extracting features using a peak visualization method (PVM). This method selects peaks with a unique width and height range to perform better classification. The vectors are divided into K random height ranges H and width ranges W of peaks-feature-set V is computed, which comprises the number of peaks, position of peaks, value of prominence of each peak, value of relative maxima, and minima. These works used deep learning but mostly focused on the temporal aspect of EEG signals. However, multi-dimensional EEG signals have meaningful features in both the spatial and temporal fields. Therefore, Singhal *et al.* [2] proposed a deep learning architecture that uses a combination of Fast Fourier Transform (FFT), Convolution Neural Network (CNN), Long Short-Term Memory (LSTM), and the recently proposed attention mechanism for extracting spatio-temporal features. The architecture is based on the approach of a sliding window from a Hankel matrix (catalecticant matrix) to get a better perspective of the data, and FFT is applied on each row for preprocessing. The CNN architecture extracts the spatial features, LSTM, and an attention mechanism extract temporal features.

Our study examines whether ensemble methods on temporal data and images could increase accuracy performance over other machine learning methods. The proposed method uses transformers in conjunction with other DNN and classical machine learning techniques trained on both time and frequency features. We propose an ensemble of transformers and bi-directional LSTM networks which take the raw signal data as input. Additionally, using two-dimensional CNNs and transformers, which are then processed into an image format. These models are combined with an XGBoost [17] model with Fourier [13], and Wavelet [12] transform features. To enhance diversity, we combine three signal processing domains: time series, 2D images, and tabular data. The proposed methodology consists of three phases. The first phase pre-processes the data and prepares the data for each domain using: Fast Fourier Transform (FFT) [18], Independent Component Analysis (ICA) [19] and Symbolic Aggregate approXimation (SAX) [20]. The second phase comprises deep learning networks for each domain: Long Short-Term Memory (LSTM) for time series, EfficientNet [21] for 2D CNN, and fully connected for tabular data. The final phase combines all the features and models into one layer for classification.

A publicly available dataset for EEG classification of alcoholic and non-alcoholic subjects was used to compare our method with the state-of-the-art EEG-NET [22] deep learning model. We perform group 5-fold cross-validation and vanilla 5-fold cross-validation. The former is used to avoid data leakage between the subjects, and the vanilla cross-validation is used for comprehension with previous works. Our method outperformed all other evaluated methods. In a group split 5-fold cross-validation, an accuracy of 85.24% and 98.06% in a random split 5-fold cross-validation.

Our main contributions can be summarized as follows:

- We present a novel ensemble method for temporal data that combines classical feature engineering, convolutional neural network, transformers, and recurrent neural networks.
- We utilized effective attention mechanisms and presented a novel technique for computing attention weights by capturing cross-dimensional interaction to improve accuracy performance.
- Our method outperforms all other methods for AUD classification using EEG.

II. Method

Ensemble learning is a meta-approach to enhancing accuracy performance by combining predictions from multiple models. Ensemble methods increase accuracy because individual models have internal biases. By combining diverse models, biases are reduced. Therefore, the method produces multimodal data (tabular features, sequential features and images) from the EEG and utilizes multimodal training. The system uses the ensemble learning technique, which enables optimal domain extraction and incorporates different models' points of view.

In our method, we have three main components as shown in Fig. 1 and explained in the following sections: (1) preprocessing of each modality - aiming to produce multimodal data, (2) specifying a model for each modality, and (3) connecting the outputs from each modality - aiming to optimize the modalities and obtain an overall prediction. A detailed description of the three components is provided below.

A. Preprocessing

Generally, EEG signals are stationary and cyclic. When a study is exposed to an image, the effect is observed after 250-350 milliseconds for each signal. We could filter and clean the signals using the following feature engineering techniques. We could also detect major changes in specific signals, recurring patterns, and correlations between signals.

1) Automatic Changepoint Detection: We can identify when and where EEG signals change using the Facebook Prophet algorithm [23]. This algorithm is used to identify and determine the magnitude of each changepoint. To detect change points, Prophet first specifies a large number of potential change points at which the rate will change. Next, it puts a sparse prior on changes in rate magnitudes (equivalent to L1 regularization) - this means Prophet has many locations where rates can change but selects as fewer of them as possible. The output of this algorithm is the location of the changepoint and the magnitude of the change. We use this output to train the LSTM model.

2) Data Granularity Reduction with PAA-SAX: First, the time series data is downsampled with Piecewise Aggregate Approximation (PAA), and then the mean values are quantized. Then, Symbolic Aggregate approXimation (SAX) [24] is applied to receive a symbolic representation of the non-discrete data. This representation enables dimensional/numerous granularity reductions, along with the definition of distance measures based on the symbolic level. Using the latter feature, one can run certain classical ML algorithms on the efficiently manipulated symbolic representation, which produces almost identical results to those of algorithms operating on the original data, yet with faster convergence.

3) Sequential Pattern Mining with Prefixspan: PrefixSpan is an algorithm for discovering sequential patterns in sequence databases [25]. By applying this algorithm to the raw EEG dataset, we could mine frequent sequential patterns that can help distinguish between alcoholic consumption studies and control studies.

4) Image Based Preprocessing: Since the dataset is multi-channel, it can be considered a two-dimensional matrix and therefore be displayed as an image. Since CNN has been proven to be SOTA on images, it is possible to assemble the same EEG image when the timestamps are on the X-axis and the channel values are on the Y-axis. Additionally, since the EEG signal is noisy, we applied the Fourier [18] Transform to a time domain signal. We converted it to the frequency domain (spectrogram) to remove the higher frequencies for a cleaner signal.

5) Data Decomposition with ICA: Multiple time series are combined by applying Independent Component Analysis (ICA) [19] to find hidden factors in a signal. The main idea is to filter and optimize signals to find the "real signal". Using Independent Component Analysis, a multivariate signal can be broken down into independent non-Gaussian components. For example, voice is usually composed of signals added

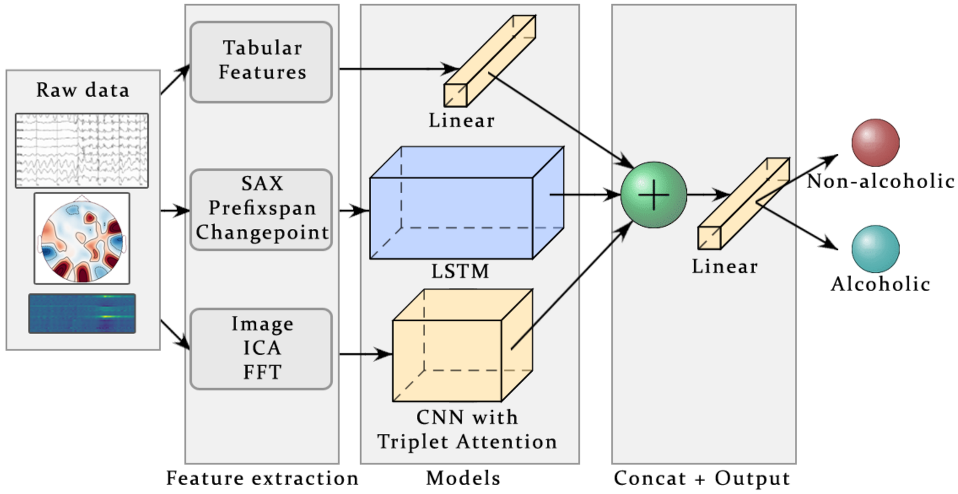


Fig. 1. Ensemble model architecture. Input: Raw data. Preprocessing: tabular, time series and image transformations. Models: Linear layers, LSTM (Figure 3), CNN (Figure 4. Concatenated and sent to a linear layer with 2 possible outputs: Alcoholic or non-alcoholic.

numerically from multiple sources at each time t. Thus, the question arises as to whether these contributing sources can be separated from the observed total signal. The ICA separation of mixed signals gives reliable results as long as the statistical independence assumption is accurate.

Using the raw data, ICA, and FFT results we generated a 3 channels image (Fig. 2).

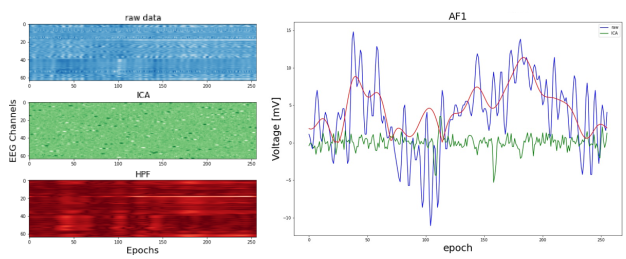


Fig. 2. The first channel (blue) describes the RAW data, the second channel (green) describes the data after ICA, and the third channel (red) describes the data after cleaning high frequencies with FFT. Raw ICA and FFT (AF1) can be seen on the right.

The EEG raw signals are noisy, and with the FFT, the signal is smoothed; using the ICA significantly lowers signal amplitude and prevents interference between signals. By transforming the signals into images, we can use computer vision techniques like CNNs and find the correlation between signals.

B. Classification Models and Algorithms

The three classification models in the ensemble model, as shown in Fig. 1, are: (1) a simple fully connected neural network, (2) an LSTM-based network (Fig. 3), and (3) a CNN for image classification (Fig. 4). Each of these neural networks was first optimized separately for the highest performance as a single model. Then a linear layer optimized the integration of the three models (The ensemble layer).

The fully connected classification model gets the EEG training data as a transposed table. The tabular data is routed through several connected layers before being forwarded as input to the ensemble layer. The LSTM model takes temporal data as input, as shown in Fig. 3. The data is forwarded through two bi-directional LSTM and two linear layers. Then, the outputs of these networks are fed to a swish activation function and concatenated to an average pooling layer as an input to the ensemble layer.

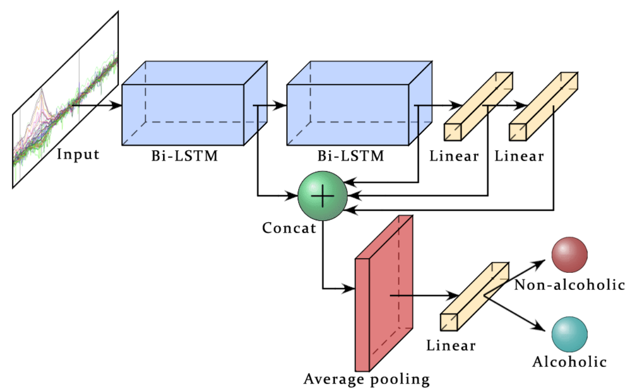


Fig. 3. The time series model architecture. The input is sent through four layers (two Bi-LSTM and two linear), and each of their outputs is concatenated and then forwarded to an average pooling layer as an input to the ensemble layer.

With CNNs having proven their efficiency in image processing previously, we use them to process the extracted images from the multi-channel time series data (the raw data, the decomposed data with ICA, and the FFT spectrogram). We used the EfficientNet [21] as a CNN classification network with attention triplet loss. The EfficientNet family of models was chosen due to their high training efficiency and relative performance. Specifically, the EfficientNet-b0 model, the smallest model architecture in the EfficientNet series, was chosen to avoid overtraining on small images and to perform faster model training.

To enhance accuracy performance we added effective attention mechanisms and present triplet attention (Fig. 4), a novel method for improving the feature representations generated by

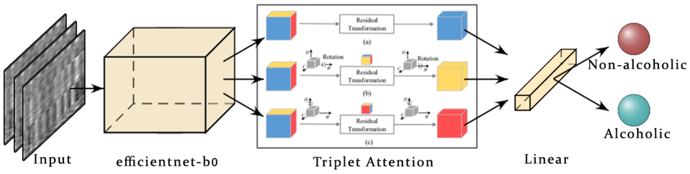


Fig. 4. The image-based input model architecture. First, the image input is sent to an EfficientNet CNN. Then, the outputs are fed into a triplet attention layer and forwarded to an average pooling layer as an input to the ensemble layer.

standard convolutional layers by explicitly building dependencies among channels or a weighted spatial mask for spatial attention. Learning attention weights allows the network to focus on target objects by learning where to pay attention. We used a three-branch structure, which focuses the network's points of interest through a heatmap [26].

III. Evaluation

This section evaluates the combination of classic pattern recognition and temporal abstraction methods with deep sequential neural networks and image-based classification approaches on EEG using CNNs. We compare the modeling approaches and their ensemble. Bringing together approaches from different domains enhances diversity and should increase accuracy. Based on the results of previous studies, we focused our study on deep learning techniques for both models (time-series and image-based). The first two approaches were designed to improve the third ensemble experiment.

A. Experimental Plan

The first two experiments aim to find the most efficient preprocessing and parameters for the ensemble model. Every experiment used the same hyper-parameters (loss, optimizer, learning rate, etc.).

1) Experiment A - Determine The Optimal Temporal Preprocessing and Feature Extraction Methods: This experiment's main objective is identifying which feature selection and preprocessing methods are optimal. As detailed in Section 2.B, we utilize a basic LSTM-based model to achieve this goal. The same parameters are used for all experiments except for the preprocessed changes. We use the AdamW optimizer, cross-entropy loss function, and cosine-annealing learning rate scheduler with 0.01 as the base learning rate for training. This experiment examines the model with three types of preprocessing: Changepoint detection, PAA-SAX, and PrefixSpan.

Changepoint detection
In order to detect changepoints, we use the Facebook prophet algorithm [23]. Since EEG signals are one-second intervals, we disable all seasonal properties. Furthermore, we also select predefined 25 optional changepoints, where the latest changepoint can occur at 850 milliseconds (15% before the end of the recorded signal). To find the minimum changepoints, a Monte Carlo Markov Chain [23] approach is used with a threshold of 2.5. The result of this method, which will be the input to the LSTM-based model, is a vector of changepoints location for each EEG signal.

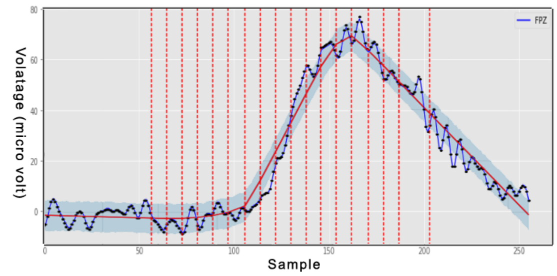


Fig. 5. Raw data of a single channel (FPZ) with the potential change point to be examined.

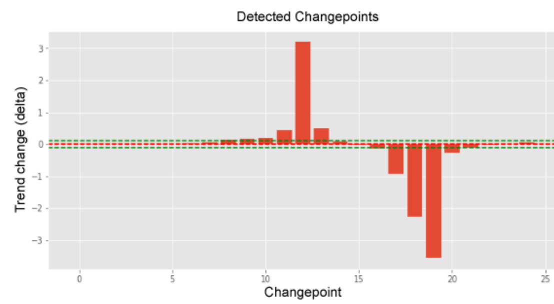


Fig. 6. The magnitude of each potential change point

Piecewise Aggregate Approximation - Symbolic Aggregate approXimation (PAA-SAX)
The SAX symbol and cutoffs are the inputs to our temporal data classification model. We applied sensitivity analysis to determine the optimal number of symbols and intervals (SAX hyper-parameters) to increase the results.

Prefixspan
Prefixspan will be applied to the training data to identify which patterns appear in the alcoholic study participant and which do not. The pattern recognition algorithm is based on the SAX-16 (SAX quantization interval width of 16) algorithm results. It determines which pattern exists for each signal, using the maximum pattern length of 128 milliseconds. The SAX-16 vectors were constructed from all the trained data using each channel signal (converting the 256 points into 16 integer points). The prefixspan was used to generate a list of existing patterns for each channel and identify which patterns exist for each SAX-16 signal channel.

Long Short-Term Memory (LSTM)
As specified in section B, the optimal preprocessing and feature extraction parameters are determined by experiments using basic LSTM. Then, we use the extracted features as the input of the LSTM-based classification model (Fig. 3). The best-performing methods will be chosen as part of the ensemble model.

2) Experiment B - Determine the Optimal Image Preprocessing and Feature Extraction Methods: Since the

EEG data is multivariate in the time domain, it can be transformed into a two-dimensional space (image). Furthermore, A channel dimension in the image can also be used to combine different types of preprocessing in each channel. We can use this approach to explore new algorithms that have yet to be applied on EEG signals. CNNs have already been proven to be the state-of-the-art algorithm for image classification (IMAGENET, CIFAR, ECT). We used the Efficientnet-b0 architecture with triplet attentions as described in section B, using the following preprocessing methods:

Independent Component Analysis (ICA)
The main goal of the ICA is to decompose the raw EEG signals into independent non-Gaussian signals. We performed several tests to find the optimal number of ICA components. The maximum number of ICA components could be 64, as the original number of EEG channels.

Fast Fourier Transform (FFT)
One of the EEG signal processing challenges is noise. We applied FFT to construct a high-pass filter and filter signals higher than 40 Hz. The filtering threshold is based on the image domain.

Two Dimensional Convolution Neural Network
In this experiment, the CNN model (as illustrated in Fig. 4) is trained with the same hyperparameter (15 epochs with a dynamic learning rate). Then, we applied the attention layer to observe which EEG channels contribute the most (and at which time point) to the final classification. Appling sensitivity analysis to this experiment could increase the accuracy performance.

3) Experiment C - Classification with Ensemble Learning: The ensemble experiment was built and designed based on the results of the experiments of each method individually. We applied all the approaches that yielded the best results for each method and integrated them within the same pipeline. First, on each model training, we performed pre-train for each model and froze the other models. Finally, we trained the whole pipeline (15 epochs) as we updated all the model weights for each step. We focused on SAX prepossessing with different intervals for the time series classifier that yielded the best results. We combined all the processing methods for the image classifier and used each of the processing as a different image input.

B. Dataset

The 'EEG Database' [27] was originally collected by Henri Begleiter at the Neurodynamics Laboratory at the State University of New York Health Center to explore the possibility of genetic predisposition to alcoholism. It contains EEG data on 122 subjects collected in 120 trials. The dataset includes data for 45 non-alcoholics and 77 alcoholics.

The trial participants were put into two distinct groups: Alcoholic and non-alcoholic. Each participant dealt with the standardized pictures from 1980s Snodgrass and Vanderwart set [28], and the EEG were recorded after they were exposed to a single or two stimuli pictures (S1 and S2, for short) and had to describe whether the S1 and S2 stimuli pictures matched or not. The EEG data were collected through 64 electrodes (Fig. 7), each sample length is 1 second, and the sample rate is 256 Hz [27].

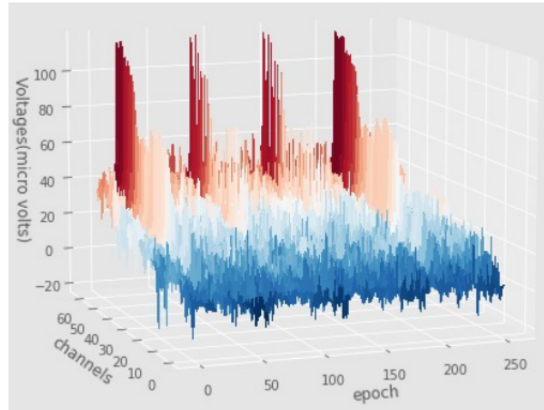


Fig. 7. 3D visualization of the EEG collected data. The x-axis represents the 64 input channels, the y-axis represents the time sample, and the z-axis represents the voltage level.

1) Cross-Validation strategy: The cross-validation strategy was grouped 5-fold by patient ID. The signal from one ID was guaranteed to be used in only one fold, ensuring no leakage between the folds and preventing over-learning based on information in the signal that is unique to the subject. There was a balanced distribution of alcohols and controls in each folder with the total number of subjects.

IV. Results

This section conducted the results of the time-series and image-based models for determining the optimal parameters for the ensemble model. The evaluation metric we used for comparing the models is the accuracy metric since the data is fairly balanced and the two classes have equal importance. Additionally, this is an intuitive metric, and similar research uses the same metric as well.

A. Experiment A - Results

The first batch of experiments was done with the raw temporal data as input using three different time series preprocessing methods: Changepoint detection, prefixspan, and SAX. Table I shows the best results of each preprocessing method.

Our research has revealed a significant difference in using preprocessing techniques in the time-series domain. Compared to SAX preprocessing, the results from changepoint detection and prefixspan are less significant. The reason is probably due to hyperparameter tuning. Additionally, there is only a minor improvement in accuracy when the SAX intervals are changed.

While splitting the data into 5,8 segments, sensitivity analysis was used to find the optimal intervals. Fig. 8 illustrates the sensitivity of the intervals - smaller intervals are more accurate. SAX was chosen with an interval of 3, which reduced the number of data points per signal from 256 to 85. As we

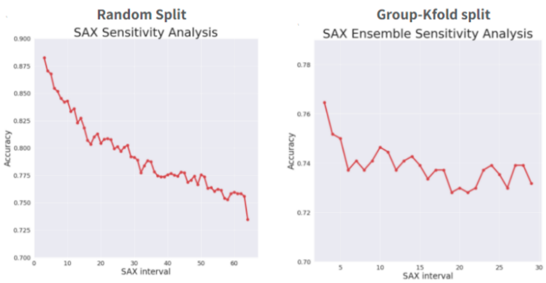


Fig. 8. The sensitivity of the SAX intervals; When the train-test split is random, there is a greater effect than using a group k-folds split.

will see later, this improved the accuracy slightly but mostly improved the convergence of the LSTM model. The LSTM model converged after 12 epochs without applying SAX to the input data and after five epochs with SAX. Additionally, SAX-16 results were used to find repeated signal types using the prefixspan algorithm.

Our implementation of changepoint detection produces less satisfying results, and accuracy is less than other methods. Fig. 5 presents the raw data of one channel from a single experiment with the potential change point to be examined. Fig. 6 demonstrates the magnitude of each potential change point. The Monte Carlo Markov Chain approach was used with a threshold of 2.5. Consequently, changepoint numbers 12 (around 375 milliseconds) and 19 (around 675 milliseconds) were detected in this example.

TABLE I
CLASSIFICATION ACCURACY WITH TIME SERIES FEATURES

Preprocess	Model	5-fold CV
Changepoint	LSTM	0.744
Prefixspan	LSTM	0.800
SAX_19	LSTM	0.811
Raw data	LSTM	0.820
SAX_3	**LSTM**	**0.828**

Ensemble of the SAX temporal feature extraction method with the LSTM-based model has achieved the highest average 5-fold cross-validation accuracy of 82.8% SAX with a time interval of 3 with the raw data, without feature extraction, as an input achieved the best results.

B. Experiment B - Results

The 2D-CNN was examined with and without the triplet attention layer. In all of our experiments, the model with triplet attention yielded better results. Fig. 9 describes an example of the triplet attention layer visualization, which shows that in both the time and the space axes, there are distinguishable signal changes when comparing alcoholic and non-alcoholic test participants.

We have found that reducing the default number of 64 ICA components to 32 has improved the results. As described in table II, using a combination of ICA and RAW data gave a better result than only raw data. With this method, we reached a 5-fold average cross-validation accuracy of 84.4%.

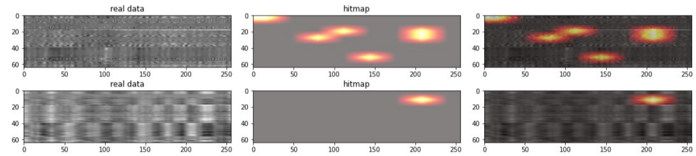


Fig. 9. Triplet attention heatmap visualization. A raw EEG as an image on the left, attention heatmaps in the middle, and an image and superimposed heatmap on the right. The non-alcoholic samples are on top, and the alcoholic samples are at the bottom.

TABLE II
CLASSIFICATION ACCURACY BASED ON IMAGE FEATURES

Preprocess	Model	5-fold CV
Raw data	B0	0.833
FFT + Raw data	B0	0.840
ICA + Raw data	**B0**	**0.844**

C. Experiment C - Results

TABLE III
EXPERIMENT 3 RESULTS - ENSEMBLE LAYER

Preprocess	Model	5-fold CV
Raw temporal/Raw image	LSTM/B0/FC	0.836
SAX/ICA/Tabular	LSTM/B0/FC	0.836
SAX/Raw/Tabular	**LSTM/B0/FC**	**0.852**

The ensemble model used three data models: temporal, tabular, and image-based. Based on the previous experiments, we chose the optimal feature extraction method and its parameters and utilized them in the ensemble model as illustrated in Fig. 1). The time-series models have demonstrated less accuracy than image-based models, as described in Table III. Overall, the ensembled model was able to refine and classify the data to the highest average 5-fold cross-validation accuracy of 85.24%.

As mentioned in section III-B.1 (Cross-Validation strategy), the results reported in this section were based on a cross-validation strategy that split the data into k-groups in a way that each test participant appears in one group. Table IV details the results of related works compared to our proposed method on the same dataset. The column 'Accuracy (ID split)' shows the experiment results on validation data that groups of participant IDs split up, whereas column 'Accuracy (random split)' shows the results on data split randomly between the training and the validation sets.

V. DISCUSSION

We examined a new ensemble method for producing multimodal data from temporal multivariate time series. We converted the EEG signals into images, produced temporal features, and ensemble these multimodal models using an ensemble layer. Analyzing the results shows that an ensemble model achieves better results than using only one method.

In examining each separate method, the results section demonstrates that pre-processing significantly impacts the model's ability to learn efficiently and achieve higher accuracy.

TABLE IV
RELATED WORK RESULT COMPARISON (*OUR IMPLEMENTATION)

Reference	Classifier	Accuracy - Random Split	Accuracy - ID Split
Kumah *et al.*	SVM and fuzzy clustering	88%	-
Shri *et al.*	BPNN and SVM	90%	-
Acharya *et al.* [7]	SVM	91.7%	-
Faust *et al.* [9]	Classic classification and neural network	92.40%	-
Zhu *et al.* [14]	SVM and KNN	95.8%	-
Fayyaz *et al.* [16]	LSTM	90%	-
Farsi *et al.* [15]	ANN and LSTM	93%	*81.19%
Singhal *et al.* [2]	CNN-LSTM-ATTN	97.5%	-
Proposed Method	SAX, ICA Ensemble-learning: CNN2D, LSTM and FC	**98.06**%	**85.24**%

The data granularity reduction by SAX mainly contributed to speeding up the convergence time, but accuracy was also slightly improved. A significant contribution to the relatively high accuracy of the model was the transformation of the temporal EEG data into images. This transformation technique, with just the raw image data and an image classification model, produced better results than our time series classification models.

With the image classification model, ICA provides the best accuracy, but the raw image data provides the highest accuracy when using the ensemble model. The time-series model using SAX-3 achieved the highest accuracy, but the ensemble model with SAX-19 achieved better results.

Other feature engineering methods, such as changepoint detection and prefixspan, did not enhance accuracy. There is still potential for improvement, and these methods have not been fully utilized. Most likely, this is due to non-optimal hyperparameter tuning.

At first glance, it may seem that similar works that were done with the same EEG dataset have reported better results than ours, for instance, 85.24% (Table III) vs. 97.5% (Singhal *et al.* [2]). However, we should distinguish between the validation results of experiments conducted with data split randomly and data that participant groups split. Data that is split randomly between the training set and the validation set will most likely result in a data leak, leading to overfitting the model. All the studies presented in the related work section and table IV split the data randomly.

Therefore, we have decided to implement one of the related works' algorithms and train it with leakage-proof validation (based on splitting the participants into diffrent groups). Overall, when we combine our two techniques, we can see an improvement in the validation accuracy compared to other research: 98.06% (Table IV) vs 93% (Farsi *et al.* [15]), when using random split, and 85.24% (Table IV) vs 81.19% (Our implementation of Farsi *et al.*, table IV), when using participant split.

Based on our results, we conclude that ensemble techniques on the Alcoholic EEG dataset are efficient and accurate compared to other related approaches.

VI. CONCLUSION

Alcoholism diagnosis with traditional tests based on blood tests and questionnaires is subjective to the patient and the examiner. The study's main purpose was to develop robust electroencephalography (EEG) classification algorithms that can produce explainable predictions that improve upon previous research results. Our study examined an ensemble technique for detecting alcoholism by examining EEG signals from three perspectives: tabular, time series, and image views. We have shown that using an image transformation technique to convert 64-channel temporal data into images using FFT and ICA improves classification accuracy significantly. Overall, the highest cross-validation classification accuracy we received was 85.5% in comparison to the state-of-the-art accuracy of 81.19%. This ensemble method can produce multimodal data (tabular features, sequential features, and images) based on any signal in any domain. Our main limitations are a narrow focus on the EEG domain and basic engineering features. Further research into changepoint detection and channel correlation algorithms and using the same ensemble method for other EEG datasets and other domains may address these limitations.

VII. CODE AVAILABILITY

The datase and our method's reproducible source code are available at https://github.com/OrKatz7/Alcoholism-Detection.

REFERENCES

[1] J. Rehm, "The risks associated with alcohol use and alcoholism," *Alcohol Research & Health*, vol. 34, no. 2, p. 135, 2011.

[2] V. Singhal, J. Mathew *et al.*, "A deep learning architecture for spatio-temporal feature extraction and alcoholism detection," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2021, pp. 1–4.

[3] P. Jahankhani, K. Revett, and V. Kodogiannis, "Data mining an eeg dataset with an emphasis on dimensionality reduction," in *2007 IEEE Symposium on Computational Intelligence and Data Mining*. IEEE, 2007, pp. 405–412.

[4] B. Porjesz and H. Begleiter, "Alcoholism and human electrophysiology," *Alcohol Research & Health*, vol. 27, no. 2, p. 153, 2003.

[5] V. Bajaj, Y. Guo, A. Sengur, S. Siuly, and O. F. Alcin, "A hybrid method based on time–frequency images for classification of alcohol and control eeg signals," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3717–3723, 2017.

[6] S. Agarwal and M. Zubair, "Classification of alcoholic and non-alcoholic eeg signals based on sliding-ssa and independent component analysis," *IEEE Sensors Journal*, 2021.

[7] U. R. Acharya, S. V. Sree, S. Chattopadhyay, and J. S. Suri, "Automated diagnosis of normal and alcoholic eeg signals," *International journal of neural systems*, vol. 22, no. 03, p. 1250011, 2012.

[8] N. T. Rachman, H. Tjandrasa, and C. Fatichah, "Alcoholism classification based on eeg data using independent component analysis (ica), wavelet de-noising and probabilistic neural network (pnn)," in *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. IEEE, 2016, pp. 17–20.

[9] O. Faust, W. Yu, and N. A. Kadri, "Computer-based identification of normal and alcoholic eeg signals using wavelet packets and energy measures," *Journal of Mechanics in Medicine and Biology*, vol. 13, no. 03, p. 1350033, 2013.

[10] C. Ekaputri, R. Widadi, and A. Rizal, "Eeg signal classification for alcoholic and non-alcoholic person using multilevel wavelet packet entropy and support vector machine," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2020, pp. 1–4.

[11] E. Malar and M. Gauthaam, "Wavelet analysis of eeg for the identification of alcoholics using probabilistic classifiers and neural networks," *International Journal of Intelligence and Sustainable Computing*, vol. 1, no. 1, pp. 3–18, 2020.

[12] A. Graps, "An introduction to wavelets," *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[13] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.

[14] G. Zhu, Y. Li, P. P. Wen, and S. Wang, "Analysis of alcoholic eeg signals based on horizontal visibility graph entropy," *Brain informatics*, vol. 1, no. 1-4, pp. 19–25, 2014.

[15] L. Farsi, S. Siuly, E. Kabir, and H. Wang, "Classification of alcoholic eeg signals using a deep learning method," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3552–3560, 2020.

[16] A. Fayyaz, M. Maqbool, and M. Saeed, "Classifying alcoholics and control patients using deep learning and peak visualization method," in *Proceedings of the 3rd International Conference on Vision, Image and Signal Processing*, 2019, pp. 1–6.

[17] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[18] H. J. Nussbaumer, "The fast fourier transform," in *Fast Fourier Transform and Convolution Algorithms*. Springer, 1981, pp. 80–111.

[19] P. Comon, "Independent component analysis," 1992.

[20] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003, pp. 2–11.

[21] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[22] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[23] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[24] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: a novel symbolic representation of time series," *Data Mining and knowledge discovery*, vol. 15, no. 2, pp. 107–144, 2007.

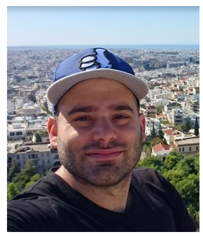
[25] J. Han, J. Pei, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth," in *proceedings of the 17th international conference on data engineering*. Citeseer, 2001, pp. 215–224.

[26] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou, "Rotate to attend: Convolutional triplet attention module," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 3139–3148.

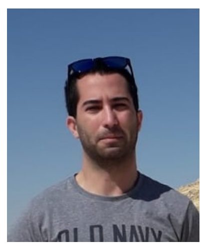
[27] H. Begleiter, "Eeg database data set," *New York, NY: Neurodynamics Laboratory, State University of New York Health Center Brooklyn*, 1995.

[28] J. G. Snodgrass and M. Vanderwart, "A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity." *Journal of experimental psychology: Human learning and memory*, vol. 6, no. 2, p. 174, 1980.

Seffi Cohen is an AI researcher. He received a B.Sc degree in computer science from MTA, Tel-Aviv in 2010, and an M.Sc degree in computer science from the Open University, Israel, in 2018. He is currently studying for a Ph.D. degree in software and information systems engineering at Ben Gurion University. He founded the first data science team in the Israeli Defense Force and led dozens of various AI solutions that were successfully deployed and yielded tangible operational value. Seffi is a Kaggle master. He reached the top of the leader board in many international ML competitions.

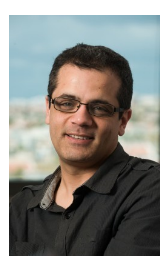


Or Katz is a computer vision and deep learning researcher. He received a B.Sc degree in electrical engineering from Afeka College of Engineering in Tel-Aviv. He is currently studying for a Master degree in software and information systems engineering at Ben Gurion University.



Dan Presil is a computer vision and deep learning researcher. He received a B.Sc degree in software engineering from Afeka College of Engineering in Tel-Aviv. He is currently studying for a Master degree in software and information systems engineering at Ben Gurion University.

Ofir Arbili is a ML researcher. He received an EMBA in Finance from The Hebrew University of Jerusalem and BA in Computer Science from Reichman University. He is currently studying for a Master degree in software and information systems engineering at Ben Gurion University.



Lior Rokach is a data scientist and a Professor of Software and Information Systems Engineering (SISE) at Ben-Gurion University of the Negev (BGU). His research interests lie in the design and analysis of Machine Learning and Data Mining algorithms and their applications in Recommender Systems, Cyber Security and Medical Informatics. Prof. Rokach has established the machine learning lab at BGU which promotes innovative adaptations of machine learning and data science methods to create the next generation of intelligent systems. Prof. Rokach is the author of over 300 peer-reviewed papers in leading journals and conference proceedings, patents, and book chapters.