Diffusion Models with Graph Attention for Spatiotemporal EEG Inpainting

by

Anindita Dutta
22101031
Angkon Dutta Joy
22101024
Plabon Mondal
22101362
Humayra Anjum Nisha
22101081

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2026.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Anindita Dutta	Angkon Dutta Joy
22101031	22101024
Plabon Mondal	Humayra Anjum Nisha
22101362	22101081

Approval

The thesis/project titled "Diffusion Models with Graph Attention for Spatiotemporal EEG Inpainting" submitted by

1. Anindita Dutta (22101031)
2. Angkon Dutta Joy (22101024)
3. Plabon Mondal (22101362)
4. Humayra Anjum Nisha (22101081)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 28, 2026.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Swakkhar Shatabda

Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Contamination of artifacts, and loss of signal is a major impairment in Electroencephalography (EEG) recording reliability and standard methods of interpolation do not resolve the multichannel brain signal complex spatial-temporal dependencies. Current state-of-the-art approaches such as transformer or GAN based methods either do not consider EEG channels interaction or have high computational resources and cannot maintain critical frequency-domain properties. The synergistic combination of diffusion models with graph attention mechanisms and low-rank decomposition of the diffusion, which can both use the learned spatial relationships and specify mathematical priors, is still an open challenge in the field. In this paper, a new model that combines Denoising Diffusion Probabilistic Models (DDPM), Graph Attention Networks (GAT), and Singular Value Decomposition (SVD) to reconstruct corrupted EEG segments with 32 channels is suggested. The DEAP dataset (32 subjects, 40 emotion-elicitation trials) goes through preprocessing processes like baselines correction, channel-wise z-score normalization, percentile clipping, and 640-sample windowing with 50% overlap. SVD estimates first 16 orthogonal spatial-temporal components, which are used as priors and fused with raw features, and a hierarchical 1D U-Net with graph attention is then trained in 300 steps to reconstruct masked parts of the EEG. The DDPM+GAT+SVD model gives a Mean Squared Error (MSE) of 0.014153 ± 0.014911, Root Mean Squared Error (RMSE) of 0.110374 ± 0.044388, Mean Absolute Error (MAE) of 0.087305 ± 0.034363, Pearson Correlation Coefficient (PCC) of 0.8766 ± 0.1207 and Power Spectral Density (PSD) distance of 0.001293 ± 0.003648, ∼80.6% RMSE improvement over DEAP DIVE baseline. Ablation experiments bear out the synergistic contributions: DDPM+SVD (MSE 0.017276 ± 0.015000) and DDPM+GAT (MSE 0.026832 ± 0.020000) are worse by 22.1 percent and 47.2 percent, respectively, which confirms that mathematical decomposition, learned spatial attention and temporal diffusion are complements. Generalization without changes in architecture is proven by cross-dataset validation on PhysioNet Motor Imagery dataset (MSE 0.196824 ± 0.461074, RMSE 0.295641 ± 0.330788, PCC 0.8387 ± 0.1478). SVDs are also pre-computed in the process which saves training time which can be deployed with ease. These findings show that diffusion models structured hierarchically by incorporating graph attention and low-rank decomposition offer robust and physiologically plausible empirical findings on EEG reconstruction and are more effective in clinical diagnostic (as well as brain-computer interface and affective computing) tasks, compared to traditional artifact removal.

Keywords: EEG Reconstruction, EEG Inpainting, Denoising Diffusion Probabilistic Models, Graph Attention Networks, Singular Value Decomposition, Artifact Removal, Spatial-Temporal Modeling, DEAP Dataset, PhysioNet, Ablation Study, Multi-Channel Signal Processing

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Electroencephalography (EEG) signal analysis forms the basis of studies on how the brain responds electrically to various stimuli, with a particular focus on emotional stimuli. Signal integrity is the fundamental ingredient of the quality of any analysis conducted on this data. These signals are however often tampered with by missing or corrupted data, which can result from electrode failure, environmental noise, or signal dropout. These interferences become a major problem for any later analysis, especially when contemplating the study of continuous neural processes. The more traditional approaches to missing data, e.g. linear interpolation or mean imputation, are generally inadequate to characterize the complex non-linear and dynamic nature of EEG signals. This weakness is especially disadvantageous in the analysis of subtle brain states, in which the timing and the shape of the signal itself contain critical data.

Recently Diffusion Probabilistic Models (DDPMs) have been a promising solution for EEG missing data reconstruction. These models function by adding noise to the data gradually and then feeding it to a neural network and letting it do the reverse, learning how to create realistic signals from noise. DDPMs capture the temporal coherence and statistical characteristics of a signal so that they can be used to fill in missing data and reconstruct the missing data with high fidelity. Standard DDPMs use, however, each channel of the EEG individually. This implies that they can miss significant links between channels. In order to address this issue, we incorporate Singular Value Decomposition (SVD) to our model. SVD decomposes the EEG signal in three components: spatial (U), importance weights (S) and temporal ($V^\top$) components. The separation enables the model to calculate the spatial and temporal patterns separately. It further minimizes noise by keeping the components that are of importance the most. Consequently, the diffusion model gets improved structured input data.

To further enhance performance, we integrate Graph Attention Networks (GAT) to capture spatial dependencies among EEG channels. SVD determines both the global spatial patterns of all channels, whereas GAT determines which channels are most similar to each other. GATs enable the model to concentrate on the most relevant channels for reconstruction, making more accurate and context-aware predictions by considering inter-channel relationships. This SVD and GAT combination develops two orientations of comprehending spatial data: SVD detects major patterns among

all the electrodes whereas GAT focuses on local linkages among neighboring channels. This method combined with the ability of DDPM to produce the temporal distributions results in the high quality reconstruction of data that maintains the overall structure and preserves the channel-specific details.

1.2 Rational of the Study or Motivation

Despite advances in machine learning and generative models, significant limitations persist in addressing missing data within EEG analysis. Current approaches often fail to account for the complex, time-sensitive, and dynamic nature of neural signals. Missing data can substantially degrade analytical quality and model performance, particularly in applications requiring high-fidelity neurophysiological information. This paper overcomes these shortcomings through the use of Diffusion Probabilistic Models (DDPMs) improved with Singular Value Decomposition (SVD) and Graph Attention Networks (GATs) to recreate missing EEG data sequences recorded during emotion response experiments. The focus on reconstructing middle signal segments is particularly critical, as these regions often contain essential neural transitions and contextual information regarding brain state progression.

Traditional reconstruction schemes combine both spatial and time information and hence cannot learn domain specific patterns effectively. This is addressed in our integration of SVD which breaks down multi-channel EEG signals into separate spatial components (inter-channel relationships), importance weights and temporal components (intra-channel dynamics). Such decomposition allows feature extraction in each specialized domain and noise is filtered by keeping only meaningful components. GATs help to supplement this method by training adaptive channel attention weights, which select the most effective channel spatial dependencies to reconstruct. Then, the DDPMs use these organized properties to create missing sections over time, on which denoising is done iteratively. This hierarchical structure, which includes decomposition through SVD, spatial attention through GAT, and temporal modeling through DDPM, is a generalized framework in the comprehension of the complex structure of EEG signals.

The proposed methodology holds substantial implications for applications requiring continuous, high-quality neurophysiological data. In clinical contexts, improved reconstruction quality enhances neurological monitoring and diagnostic capabilities. For brain-computer interfaces and neurofeedback systems, reduced data corruption enables more reliable real-time processing. Furthermore, human-computer interaction systems that adapt to users' cognitive and affective states benefit from the enhanced signal fidelity. By systematically addressing the persistent challenge of missing data through this integrated approach, this research strengthens the foundation for robust EEG-based analysis and its downstream applications.

1.3 Problem Statement

In EEG signal analysis, masked regions arising from electrode failures, environmental noise, or signal dropout can severely compromise the accuracy and reliability of subsequent analyses. Traditional imputation techniques, such as interpolation, are insufficient for capturing the complex non-linear dynamics of EEG signals, par-

ticularly when critical neural events are obscured. This study focuses on reconstructing masked portions of EEG data while preserving temporal continuity and cross-channel relationships. Existing approaches struggle to comprehensively reconstruct these segments without distorting the underlying signal structure. Also, the traditional approaches usually are not applicable in capturing the inherent coupling of inter-channel dependencies and signal evolution over time constraints, constraining their ability of reconstruction.
We propose an integrated approach combining Denoising Diffusion Probabilistic Models (DDPM), Singular Value Decomposition (SVD), and Graph Attention Networks (GAT). SVD breaks multi-channel EEG signals down into orthogonal spatial and temporal components that allows domain-specific processing and also removes noise by selecting components. DDPMs leverage these decomposed features for iterative refinement of missing segments, contrary to SVD that models global cross-channel associations, GATs model complementary local attention. Such integration is hierarchical and allows some form of precise contextual reconstruction. The research aims to provide a robust solution for masked EEG data reconstruction, thereby enhancing the quality and reliability of EEG analysis in applications such as neurofeedback, clinical monitoring, and human-computer interaction.

1.4 Objective

The purpose of this study is to solve the problems of missing and corrupted data in multi-channel EEG recordings, specifically in experimental setups. By combining the recent developments in generative modeling and graph-based learning, the work aims at developing and testing a sound reconstruction framework that maintains a temporal dynamics and the spatial dependencies of EEG signals. The specific objectives that this research aims to achieve are outlined as follows:

- To develop a technique using Denoising Diffusion Probabilistic Models (DDPMs) combined with Graph Attention Networks (GAT) and Singular Value Decomposition (SVD) to reconstruct missing data in multi-channel EEG signals, preserving temporal continuity and spatial dependencies while incorporating global spatial–temporal structure through SVD.

- To evaluate the impact of EEG reconstruction using the DDPM + GAT + SVD framework on the statistical properties and structural integrity of EEG signals, ensuring reconstructed segments retain characteristics of the original data in both temporal patterns and inter-channel relationships.

1.5 Research Outcome

The study examines the incorporation of Denoising Diffusion Probabilistic Models (DDPMs) and Graph Attention Networks (GATs) and Singular Value Decomposition (SVD) in reconstructions of missing parts in multi-channel electroencephalographic (EEG) signals. The proposed DDPM + GAT + SVD architecture was tested on the DEAP dataset of 32 EEG subjects consisting of a collection of 29,760 EEG segments, and was trained to reconstruct 77 timepoints in a constant collection of timeframes that masked about 70 percent of 32 EEG channels at any one time.

The architecture has a Mean Squared Error (MSE) almost 0.014153 ± 0.014911), Root Mean Squared Error (RMSE) of 0.110374 ± 0.044388), Mean Absolute Error (MAE) of 0.087305 ± 0.034363), Pearson Correlation Coefficient (PCC) of 0.8766 ± 0.1207, and Power Spectral Density distance of 0.001293 ± 0.003648 which proves that the proposed architecture beats DDPM+SVD (without GAT), by 18.1 percent on MSE reduction (0.017276 vs. 0.014153), and also beats DDPM+GAT (without SVD) by 47.2 percent on MSE reduction (0.026832 vs. 0.014153). This proves that mathematical decomposition (SVD) is better plus learned spatial attention (GAT) complement each other to their optimal performance SVD integration offers dimensional reduction to 16 major components and it lowers the training time. The architecture makes use of an iterative denoising process that uses 300 steps that include precomputed SVD components that are propagated through specific convolutional modules and GAT mechanisms (4 heads, 64 hidden dimensions) propagates information between intact and corrupted channels dynamically. This paper indicates that through structural hierarchical integration of diffusion based time modeling, graph based spatial attention, and SVD based structural priors high fidelity reconstructions can be achieved. This approach can improve the reliability of EEG based analyses in clinical neuroscience, affective computing, brain-computer interfaces and human computer interaction by reducing the effect of temporal gaps in brain recordings.

1.6 Methodology in Brief

In this paper, the proposed study will develop an improved EEG reconstruction model that combines Denoising Diffusion Probabilistic Models (DDPMs), Graph Attention Networks (GATs), and Singular Value Decomposition (SVD) to reconstruct lost parts of multi-channel EEG signals. In order to model realistic processes of data loss, contiguous masked areas around 35% of the signal length are placed within the temporal interval spanning timesteps 180–400, and the probability of masking each channel was 70%. Preprocessing EEG signals involves baseline correction on 384 samples (the first 3 seconds) and then z-score normalization channel-by-channel and clipping the amplitude to the 1st and the 99th percentiles. The information is divided into 640-sample windows that overlap 50% of the information of the DEAP dataset (32 participants, 40 trials each). The resulting dataset will be split into 80% training and 20% test sets by sampling by participants. The reconstruction framework is developed based on hierarchical 1D U-Net architecture with 96 base channels and channel multipliers . In every encoder block, there are three residual blocks having GroupNorm (8 groups) with SiLU activations. Dimension 96 sinusoidal timestep embeddings are transformed to dimension 384 through a Multi-layer Perceptron (MLP) and fed into every residual block to conditionalize the denoising procedure. The adaptive inter-channel relationship has been incorporated with multi-head Graph Attention Networks (4 heads, 64 hidden units) at encoder levels. In order to add global signal structure, SVD is used to identify the highest 16 orthogonal spatiotemporal components in the signal. The spatial components (U) and temporal components (V^{T}) are measured by distinct convolutional networks, and singular values (S) are used to dynamically composite their responses via a learnt MLP. The results of these SVD-based features are combined with original signal representations in a learnable combination mechanism, which offers low-rank

structural information across all the steps of diffusion. In the DDPM, 1000 denoising steps are used to learn temporal dynamics with a linear noise schedule (β 0.0001 to 0.02) and progressively repair masked parts. The training is done over 50 epochs with the Adam optimizer with a learning rate of 1×10^{-4}, a dropout rate of 0.05, and gradient clipping with a maximum gradient of 1.0. The performance of reconstruction is only analyzed in the masked area by the use of five criteria: Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as distances of reconstruction performance, Mean Absolute Error (MAE), Pearson Correlation Coefficient (PCC), and Power Spectral Density (PSD) distance. This integrated pipeline is an enhanced combination of preprocessing, SVD-based structural decomposition, graph-based spatial modeling, and probabilistic temporal stability in resolving the complicated spatiotemporal interconnections of EEG signals.

1.7 Scopes and Challenges

1.7.1 Scopes of the Research

- **Better Data Quality:** This study will substantially increase the completeness and quality of EEG data, thus providing better and more trustworthy data to any analytical pipeline.
- **Improved Signal Analysis:** The very high-fidelity reconstructed EEG data will enable more suitable inferences to be made by analysis models, especially in signals collected during complex cognitive or emotional tasks.
- **Wider Implications:** The results can be used to improve data pre-processing across fields such as clinical neuroscience (e.g., patient monitoring) and adaptive human-computer interaction systems.
- **Data Imputation:** Data imputation using DDPMs with SVD and GAT will contribute to making analytical pipelines more robust, less sensitive to missing data, and more practical for real-world applications.

1.7.2 Challenges of the Research

- **Data Availability:** High-quality and large-scale EEG datasets are limited in size, and often provide too few trials to effectively train a generative model.
- **Complexity of Model:** DDPMs merged with SVD and GAT in data preprocessing are computationally intensive, and their training can be both time-consuming and resource-demanding.
- **Managing Temporal Coherence:** Ensuring temporal integrity and phase continuity in generated data segments is technically challenging.
- **Generalization Across Subjects:** High inter-subject variability in EEG signals makes it difficult to build models that generalize well across individuals.
- **Real-World Applicability:** A major challenge is ensuring that the generated data is not only statistically consistent but also realistic enough for use in sensitive, real-world research and application contexts.

1.8 Thesis Organization

This thesis is divided into six chapters. Chapter 1 provides the background of the research, motivation, problem statement, objectives, expected results, methodology overall, and research scope. Existing literature on the reconstruction process of EEG, diffusion model, graph attention network, and research gaps are reviewed in Chapter 2. Chapter 3 talks about technical requirements, societal and environmental impacts, ethical considerations, requirements compliance, project management, and economic analysis. Chapter 4 presents the complete methodology including the preprocessing of the data, DDPM framework, the architecture of hybrid U-Net+GAT+SVD, training procedures and implementation details. Results from the experimental sequences are analyzed by 5 metrics (MSE, RMSE, MAE, PCC, PSD) and results of 3 model variants are compared by ablation studies, PhysioNet dataset validation, and limitations and advantages are discussed in this chapter. Chapter 6 concludes with some key findings (80.6% improvement over baseline), some contributions to the field of biomedical signal processing and recommendations for some future work including clinical deployment and optimization in real time.

Chapter 2

Literature Review

2.1 Preliminaries

Electroencephalography (EEG) is a non-invasive technique of measuring electrical activity of the brain. EEG has been used in almost all fields of cognitive neuroscience, clinical diagnosis, and affective computing, and provides high temporal resolution, which makes it appropriate to study dynamic brain processes like attention, emotion, perception, and motor control. The method normally requires the insertion of several electrodes at the scalp to record voltage changes based on neural oscillations in discrete frequency bands (e.g., delta, theta, alpha, beta, gamma). These fluctuations include valuable information related to cognitive conditions, the mental workload, and emotional reactions. EEG signals are, however, inherently noisy and prone to artifacts due to eye blinks, muscle artifacts and environmental noise. Additionally, the limited spatial resolution and high variability across subjects and sessions pose challenges for generaliz ability and interpretation. Large-scale EEG datasets are scarce, and many existing datasets suffer from incomplete recordings, either due to technical issues such as electrode detachment or due to intentional artifact rejection during preprocessing. These missing segments significantly impact downstream analysis, particularly in applications where continuity of signal is essential, such as emotion recognition or event-related potential (ERP) studies. EEG recordings typically have large missing portions, as 30-40 percent of data may be lost due to lose electrodes, equipment issues or the exclusion of noisy segments. These problematic sections are typically discarded as part of the traditional preprocessing, leading to a waste of useful information and result deterioration. Standard imputation methods (e.g., interpolation, basic modeling) are unable to address the rich non-linear structure of brain rhythms and couplings between channels. Consequently, the reconstructed data are corrupted and unreliable because they cannot be applied directly to research or clinical applications. Since the body of knowledge based on EEG-based studies finds its way more and more into the real world (e.g. brain-computer interfaces (BCIs), neurofeedback, and cognitive state monitoring), the integrity and continuity of the EEG recordings have become a topic of concern.

2.2 Review of Existing Research

Diffusion models have become a breakthrough family of deep generative models, with state-of-the-art performance in video, molecule, and image-to-image transla-

tion (Yang et al., 2023) [8]. With a unifying theme of a perturbative and reversal mechanism, and with variants including Denoising Diffusion Probabilistic Models (DDPMs), Score-Based Generative Models (SGMs), and stochastic differential equations (Score SDEs) in common, such a family of models generate high-fidelity samples.With success in beating GANs in both out-translating GANs in image-to-image translation and generating text-to-images in modalities including DALL-E 2 and Imagen, measured through performance metrics including likelihood estimation, Inception Score (IS), and Fréchet Inception Distance (FID), consistently outperforming baselines over a range of datasets including CIFAR-10 and ImageNet for translating images, and both NLP and molecular datasets for molecular and drugs, with such success stories, one such standout issue is computational efficiency and scalability, and enormous computational requirements for training and inference. But with ongoing work in improving efficiency in sampling and estimating likelihoods, such areas have a long distance yet to cover for development, and little theoretical work in diffusion processes and model performance, but a direction for deeper work, with high success in producing high-fidelity, real-appearing output, computational requirements will make and break larger use, but with integration with complementary frameworks such as integration with GANs, SGMs, and DDPMs, and extending them to new structures and domains, a new universe can possibly unfold, not only discovering strengths and diversity in diffusion models, but gaps and future work avenues. In its survey, by refining sampling techniques, developing greater scalability, and researching hybrid architectures, diffusion models can build upon current breakthroughs in generative modeling and have a deep impact for a larger range of challenging cases.

Adib et al. (2023) [4] explore the problem of generating realistro and synthetically generating ECG (electrocardiogram) signals when training medical diagnostics with machine algorithms, with regard to the fact that current ECG datasets have an unbalanced ratio of heartbeats with no disease (normal heartbeats predominate them). In an attempt to correct such a problem, a problem, the authors introduce using the Improved Denoising Diffusion Probabilistic Model (DDPM) in generating synthetic ECG information and comparing its performance with using Wasserstein GAN with Gradient Penalty (WGAN-GP), whose performance in generating real datasets is prestigious. In its approaches, the article employs a 1D ECG times seriesdata to 3-channel 2D ECG representations via Gramian Angular Summation/Difference Fields (GASF/GADF) and Markov Transition Fields (MTF), and DDPM is trained in these 3-channel 2D representations, with generated 1D ECG information generated and checked for evaluation subsequently. Precision Score, Precision-Recall Curriculum with its Area Under Curve (AUC), and AUC of Receiver Operating Characteristic (ROC) Curriculum's curve form the performance measures in checking generated synthesized information for suitability and reality in generating high-quality ECG information equivalent to real ECG information. With its study, it is proven that despite its suitability, DDPM is outperformed in performance in all performance evaluations in generating information with a relatively high and real ECG information level, suggesting its high suitability in generating high-quality and real ECG information over DDPM, whose performance in generating synthesized ECG information pales in comparison with WGAN-GP, in suggesting that its performance can possibly be upgraded with improvement in its model's structure or training

processes, future work can involve extending DDPM in generating synthetics for a larger variety of arrhythmias using conditional models and with alternative techniques for datature and training over a larger training corpus. In addition, whether generated synthetics have medical utility in real-life medical diagnostics can serve as a direction for future investigation.

Zhang et al. (2024) [23] tackled a tough problem: getting solid, subject-independent brain features from motor imagery EEG data. Old methods often mess up because there isn't enough labelled data, and people's brains differ a lot. In order to overcome these challenges, the authors developed MAE-MI, a masked autoencoder model that was designed with self-supervised learning on EEG signals in mind. The major innovation is the Connectivity Guided Masking (CGM) approach that utilizes a depth-first search to mask connected areas in the timechannel EEG grid to reduce information leakage, and force the model to reconstitute masked regions with the help of salient contextual features. The architecture uses an asymmetric transformer architecture that has eight encoder and four decoder attention blocks, optimized to the spatiotemporal characteristic of EEG data. PhysioNet EEG Motor Movement/Imagery Dataset was used to test MAE-MI in which participants were timed to perform 4 MI tasks (left hand, right hand, both fists, both feet) each sample containing two seconds of EEG data at 64 channels. The experimental pipeline included three steps: Pre-training on unlabelled data of 20 subjects, standardisation through fine tuning of a classification head on partially labelled data and individualised further fine tuning. The main measure of evaluation was taken to be classification accuracy. MAE-MI was able to reach an average single-subject accuracy of 78.2% (standard deviation = 3.7%), outperforming all the baseline models, including supervised convolutional neural networks (EEGNet, ShallowNet, DeepNet), transformer models (ViT-B, ViT-L), and unadulterated MAE. The importance of the CGM strategy and spatial design of the model was supported through ablation experiments, where either of these two elements was absent and induced a drastic decrease in accuracy. Experiments on visualisation also supported the capability of MAE-MI to reconstruct salient EEG features despite up to 90 percent masking, thus indicating excellent cross-subject generalisation. The results suggest that self-supervised pre-training of MAE-MI with the CGM methodology enables the extraction of generalisable EEG features, which is beneficial to MI recognition and offers opportunities to extend the results to wider contexts, like sleep-stage detection, denoising, and synthetic EEG. Secondly, the authors propose future studies to further refine pre-training paradigms and apply the model to other noisy or incomplete EEG tasks with the expectation that MAE-MI will allow more general and transferable EEG analysis in the field of brain-computer interface studies.

The article by Zhou and Liu (2024) [25] focused on the issue of efficient and generalisable representation learning to learn how to estimate gaze using EEG, which they pointed out resulted in most of the currently available models requiring vast amounts of training and having poor generalisation. They proposed a masked autoencoder (MAE) model that uses the self-supervised pre-training mechanism to enable the learning of strong EEG characteristics in deep learning models. The MAE system is made up of an encoder of the EEGViT hybrid Vision Transformer, which is initialised with ImageNet weights, and a decoder made up of Transformer.

In pre-training, a random selection of the EEG signal matrix is masked, the model is trained to decode the masked signals and thus forced to reproduce missing signals, forcing the encoder to learn important latent signals. The EEGEyeNet dataset used in the investigation consisted of 128-channel EEG and eye-tracking data of 27 participants (21,464 samples in total). The paper was centered on the Large Grid Paradigm, where individuals were obsessed with 25 positions on the screen and through the model, XY gaze coordinates were predicted. The data were divided into three parts training, validation and test (70, 15, 15). Root mean squared error (RMSE) of actual and predicted gaze positions in millimetres was used as the main measure of evaluation. The findings proved that MAE pre-training increased the prediction accuracy and training performance. The best model, with a decoder of 2 Transformer blocks and a masking ratio of 40-50 percent, had an average RMSE of 53.5 mm which exceeded a baseline EEGViT of 55.9 mm. The pretrained encoder achieved similar or even better performance using only a third to half the amount of training epochs as models trained at random. Fine-tuning was stabilised using more complex decoders, and best masking ratios gave even better results. It was found that it exhibited mild over-fitting but this was countered by using a more advanced decoder. The authors suggest that MAE pre-training is a promising, generalisable method of learning EEG representations and advise future studies to investigate other encoder architectures and larger and more diverse EEG datasets.

Mohammadi Foumani et al. (2024) [15] examined the underlying issues of learning semantic EEG representation in self-supervised learning, which is low signal-noise, large amplitude variation, and lack of explicit segmentation in EEG sequences that often limit the semantic fidelity and strength of the learned features. To overcome these shortcomings, the authors proposed EEG2Rep, a framework of self-prediction, which predicts masked inputs in the space of the latent representation, instead of the raw EEG domain. This approach is supported by a new Semantic Subsequence Preserving (SSP) masking scheme, where informative, contiguous sub-sequences are preserved during masking and the model is directed to richer semantic representations and trivial interpolation is avoided. The architecture of EEG2Rep consists of a three-layer convolutional input embedding, transformer-based target and context encoders and a cross-attention predictor network. The model is trained by using a combination of a reconstruction loss in the representation space and a regularization known as VICReg to avoid representation collapse and to encourage diversity. There were six heterogeneous EEG Datasets, DREAMER (emotion detection), STEW (mental workload), Crowdsourced (eyes open/closed), DriverDistraction, TUAB (abnormal EEG) and TUEV (event detection) and data were collected with Emotiv wireless headsets and clinical EEG caps. Measures of evaluation included accuracy, AUROC and weighted F1, which were measured by linear probing and fine-tuning protocols. EEG2Rep was the best in all tasks in comparison to state-of-the-art invariance-based and self-prediction baselines (e.g. TS-TCC, TF-C, BIOT, BENDR, MAEEG) and obtained the highest mean accuracy and strong resistance to noise (Gaussian, DC-shift, amplitude perturbations). The effectiveness of SSP masking and representation-space prediction was confirmed in ablation studies, providing the best results in case 50 per cent of the input was retained across three blocks. Cross domain experimentation showed that related dataset pretraining increased generalization even more. The authors emphasize that EEG2Rep

can be effective in learning robust and generalizable EEG representations and suggest further studies to make masking strategies consistent with neurophysiological time scales and to pre-train on broader EEG domains.

Wei et al. (2025) [19] examined the problem of using large-scale (unlabeled) high-density electroencephalography (EEG) data to improve model performance in cases where only a small amount of labeled low-density EEG data is known as a common bottleneck in the clinical diagnosis of brain disorders. The approaches of the past are incapable of fully utilizing the unlabeled data or moving the knowledge between high and low-density (HD and LD) EEG leading to insufficient generalization and accuracy. The authors introduced EEG-DisGCMAE, a single and pre-trained graph contrastive masked autoencoder distiller, which combines both graph contrastive, and masked autoencoder pre-training to enable effective graph-based EEG representations learning and knowledge distillation. EEG-DisGCMAE creates EEG graphs using the power spectral density characteristics of the alpha band, mainly, as well as Pearson correlation to establish connectivity, thus describing each electrode in the form of a node. The framework is a hybrid of graph contrastive pre-training (GCL-PT) and graph masked autoencoder pre-training (GMAE-PT), which allows the model to reconstruct masked graph samples and contrast the reconstructions between each other to supervise each other. A new Graph Topology Distillation (GTD) loss allows a lightweight student model to be learned on LD EEG and then learned using a teacher model on HD EEG, and thus, dealing with missing electrodes and compressing model parameters. Spectral-based graph convolutional networks (DGCNN) and spatial-based Graph Transformers can be used as backbone models. Two clinical data sets (EMBARC, 308 participants, 64 electrodes, and HBN, 1,594 participants and 128 electrodes) were experimentally evaluated, and binary classification tasks included sex, depression severity, major depressive disorder (MDD), and autism spectrum disorder (ASD). The pre-training corpus was 24k or so EEG graph samples, and downstream tasks used ten-fold cross-validation. Performance measures were the area under the curve in the receiver operating characteristic (AUROC), and classification accuracy. EEG -DisGCMAE has the best performance among all baselines with up to 87.8 percent AUROC and 86.9 percent accuracy on HBN MDD classification using large Graph Transformer models and robustness to noise and electrode dropout. The efficacy of the unified pre-training paradigm, the GTD loss and utilization of alpha band features was verified in ablation studies. The model also generalized to emotion recognition tasks on the SEED dataset and also high performance under very low-density conditions with only eight electrodes.Future studies need to focus on the heterogeneity of data between EEG acquisition systems and endeavors to create large scale, graph based, EEG foundation models.

Li et al. (2025) [28] presented CoMET, an architecture based on contrastive masking of the brain foundation, aimed at solving the shortage of universal and transferable representation of EEG analysis across different tasks, across the various dataset and at varying recording conditions. The current models of EEG are task-specific and because of the differences in the layouts of electrodes, sampling rates, and signal characteristics depending on the subject, they are poorly generalizable. In order to get through these constraints, the authors established a large-scale self-

learning environment where they learn task-agnostic EEG representations that can be transferred to the heterogeneous EEG applications. The CoMET model that is proposed is a combination of both masked signal modeling and contrastive learning to ensure local but global consistency between EEG signals and semantics. In pre-training, chunks of multichannel EEG parameters are selected randomly and masks them and recovered, and contrastive goals impose a consistency among the various augmented representations of the identical EEG segment. It has the backbone architecture based on the Transformer that allows modeling of temporal long range and scaling on datasets. Training consists of two objectives: MSE reconstruction loss on masked tokens and InfoNCE contrastive loss on global tokens of the main encoder and a momentum encoder (EMA-updated). CoMET has been pre-trained using mixed datasets across 3000+ subjects, and more than a million EEG samples, and linear probed on 10 downstream datasets of MI, ERP/P300, imagined speech, emotion recognition, SSVEP, and clinical EEG. The balanced accuracy (and also kappa/F1 on key tasks), with CoMET-Large (151M) reporting strong improvement (e.g., 62.75% $\pm$ 1.62 on BCIC-IV-2A and 73.22% $\pm$ 0.81 on BCIC-IV-2B), is used to report performance, and achieves better results on most datasets overall, which is attributed to minor weaker results on TUH-derived datasets because of overlapping baseline pretrains.

Tosato et al. (2023) [7] address key limitations in EEG-based Brain-Computer Interface (BCI) systems, particularly the scarcity and subject-specific nature of training data, by proposing a synthetic data generation framework using Denoising Diffusion Probabilistic Models (DDPM). EEG signals are first transformed into electrode-frequency distribution maps (EFDMs) using short-time Fourier transform (STFT), preserving spatial and spectral information. A DDPM based on OpenAI's improved diffusion architecture is then trained to denoise EFDMs and generate emotionally labeled synthetic EEG data. The model is evaluated on the SEED-V dataset, which contains EEG recordings from 16 subjects across three sessions annotated with emotional states (happy or sad). Classifiers trained on both real and synthetic data consistently outperformed those trained on real data alone, achieving a top accuracy of 92.98% compared to 91.43%, with performance stability confirmed over 20 training runs. The model architecture includes convolutional layers followed by a GELU-activated dense layer, optimized via cross-entropy loss. Importantly, the study avoids overfitting by ensuring synthetic data differs from training inputs. Future work includes enhancing EFDM processing with tensor inputs to improve scalability, enabling patient-specific fine-tuning with limited samples, and supporting privacy-preserving data augmentation in open-source BCI applications.

Aristimunha et al. (2023) [5] address the limitations in EEG-based sleep research due to the lack of high-quality, annotated polysomnography data and privacy regulations that restrict data sharing. To overcome these issues, the authors introduce a Latent Diffusion Model (LDM) applied to two large EEG sleep datasets—Sleep-EDFx and the Sleep Heart Health Study (SHHS)—to synthesize 30-second EEG segments representative of different sleep stages. Previous generative approaches, such as GANs, have faced challenges including training instability and mode collapse. Furthermore, earlier works on diffusion models in EEG were limited to small-scale datasets and focused on BCI applications. In this study, EEG windows are first encoded using an

Autoencoder with Kullback-Leibler divergence (AE-KL), ensuring that the learned representations capture critical neural oscillations across delta, theta, and alpha bands. These latent vectors are then input into a Denoising Diffusion Probabilistic Model (DDPM), which reconstructs the EEG data using an attention-based residual U-Net during the reverse denoising process. Evaluation of the generated signals involves metrics such as Frechet Inception Distance (FID), Multi-Scale Structural Similarity Index (MS-SSIM), and Power Spectral Density (PSD) analysis. The inclusion of spectral loss significantly improves fidelity: FID scores decrease from 11.933 to 0.308 on Sleep-EDFx and from 0.936 to 0.168 on SHHS. MS-SSIM scores exceed 0.90, and PSD comparisons show strong alignment between real and synthetic EEG, particularly in the delta band. The results demonstrate that incorporating frequency-domain structure via spectral loss enhances both the realism and diversity of synthetic EEG signals. Future directions include personalizing LDM-generated EEG based on patient-specific profiles and exploring signal-domain diffusion models that bypass latent encoding. This method offers promising solutions for expanding datasets, preserving privacy, and improving model performance in sleep-related EEG research.

Sharma et al. (2024) [6] address EEG data scarcity in clinical classification tasks by shifting augmentation to the latent space through a class-conditioned diffusion model (MEDiC). Unlike conventional time-domain augmenters and Generative Adversarial Networks (GANs), this method first utilizes a pre-trained EEGNet encoder to transform 30-second resting-state EEG windows into 784-dimensional latent vectors while preserving class-relevant features. A Denoising Diffusion Probabilistic Model (DDPM) then synthesizes new embeddings conditioned on class labels, employing classifier-free guidance and a U-Net-based noise estimator. The network integrates class and timestep embeddings through multilayer perceptrons (MLPs). The model was trained for 200 epochs using the Adam optimizer with a learning rate of 1×10^{-4} on a 60/40 split of a 65-subject dataset (36 AD, 29 CN), with an additional 23 FTD patients incorporated for 3-class generalization. To assess synthetic sample fidelity, the authors trained a separate MLP classifier on generated embeddings and evaluated it on held-out real samples. Furthermore, the Jensen–Shannon divergence score and comparison to a Variational Autoencoder (VAE) baseline were used to quantify performance. The results show a near-perfect AD vs CN class.

Fusion Models Klein et al. (2024) [13] address the problem of limited data in ERP-based EEG brain-computer interfaces by introducing a conditional diffusion model that employs classifier-free guidance to generate EEG signals tailored to individual subjects, sessions, and stimulus classes. Unlike earlier approaches that use indirect representations or generate unconditional samples, their model directly synthesizes EEG time series, improving specificity and utility. Utilizing data from the ERP study by Lee et al. (2019), which includes over 430,000 one-second EEG segments from 54 participants across two sessions, the model builds on the VP-SDE framework and integrates elements of EEG Wave and diff-EEG architectures. For evaluation, in addition to standard metrics like Fréchet Inception Distance (FID) and Sliced Wasserstein Distance (SWD), domain-specific metrics are introduced: Peak Amplitude Difference (PAD), Peak Latency Difference (PLD), and Scaled Difference in Mahalanobis Distance (SD-MD), calculated per subject-session-class. The best-

performing model, after 600k training steps, achieved an average balanced accuracy (ABA) of 0.817 on real test data, nearly matching the 0.818 of a baseline trained on actual recordings. The synthetic signals exhibited lower PAD (0.48 µV vs. 0.83 µV), PLD (0.016 ms vs. 0.042 ms), and SD-MD (2.39 vs. 5.16), demonstrating strong signal fidelity and variation. Although PLD showed minor challenges with closely spaced peaks, the model reliably captured ERP features, outperforming inter-session comparisons in FID. This conditional diffusion framework can potentially address class imbalance, enhance classifier training, and support transfer learning in EEG applications. As the first ERP-focused EEG generator conditioned on subject, session, and class, the study highlights the adaptability and interpretability of conditional diffusion for generating realistic EEG data.

Torma and Szegletes (2025) [32] explore the question of whether diffusion probabilistic models (DPMs) can be enhanced to create naturalistic EEGs and effective data augmentation, the obvious drawback is that deep EEG decoders are not only limited by a limited number of high-quality records but are also expensive and time-intensive in terms of computational efficiency. The authors use continuous-time DPMs with two acceleration strategies (denoising diffusion implicit sampling (DDIM), fewer sampling steps and progressive distillation, teacher-student halving) to be able to synthesize a single-step EEG, which is an unavoidable quality-speed trade-off. It is based on their implementation of a variant of the U-Net (sinusoidal step embeddings, two residual blocks per resolution, class conditioning via one-hot labels) and is trained on two test sets: VEPESS (18 subjects, 64 channel EEG, 512 Hz, imbalanced oddball VEP epochs ≈12.5% target; 1–40 Hz filtering, 1 s epochs) and BCIC-IV-2a (9 subjects, 22 channels, 250 Hz 4-class motor imagery; 4–38 Hz filtering, baseline correction, 4 s epochs) both split in a cross-subject manner across five folds. The tests evaluated are FID/IS (calculated with EEGNet features), SWD, electrode-correlation KS tests, and time/frequency connectivity analysis, and downstream augmentation impact EEGNet, chronoNet, SVM and kNN (based on accuracy, AUC, precision, and recall with confidence intervals). Research indicates that 1024-step models generate better quality samples, whereas the distilled one-step models are still useful and usually enhance cross-subject classification, particularly with the SVM/kNN, which leads to the idea of DPM-based augmentation as a viable direction to move to.

The paper of Chen et al. (2023) [9] deals with fewer data in diagnosing Alzheimer's Disease (AD) with Electroencephalography (EEG with a new scheme for fewer-data augmentation, Diff-EEG, for improvement in deep model performance. Overfitting, a problem with fewer EEG data, reduces effectiveness in diagnosing AD with model effectiveness, and in an attempt to counteract, a new scheme combining Vector Quantized Variational Autoencoders (VQ-VAE) and guided diffusion models is proposed for distribution learning of EEG and high-fidelity synthetic data creation. VQ-VAE is utilized for feature extraction of multi-channel EEG signals and converting them into a low-dimensional latent space, and guided diffusion models for generating high-fidelity EEG with specific labels, with high-fidelity augmentation. Performance is evaluated with accuracy, precision, and recall, with 89.3% accuracy for differentiation between subjects with and without Alzheimer's disease and NC subjects with a 2-data scale and 93.3% with a 3-data scale. An open-source resting

state-closed eyes recording with 88 subjects, including 36 with Alzheimer's disease (AD), 23 with frontotemporal dementia (FTD), and 29 with normal controls (NC), is utilized in the study. There are 19 scalp electrodes and 2 reference electrodes with a 500 Hz sampling and 10 μV/mm resolution in recordings, and recordings vary in duration with groups, with approximately 485.5 minutes in Alzheimer's disease group. Despite such positive performance, generalizability in populations and settings and model use with specific EEG features could restrict its use in neurological disease, according to the study. Authors hope future studies can extend the model to other brain disease and include additional EEG markers for even increased accuracy in diagnostics. In general, the Diff-EEG method is full of potential for enriching Alzheimer's diagnostics with enriched information in EEG, and for specifying regions for future development and work.

Siddhad et al. (2024) [17] deal with the challenge of insufficient data in EEG-based emotion recognition by proposing a Conditional Denoising Diffusion Probabilistic Model (CD-DPM) to generate realistic synthetic EEG signals while preserving class-specific emotional features. Earlier methods, such as Generative Adversarial Networks (GANs) and simple data augmentation, were not enough to produce stable signals and broad, representative datasets. Instead, their approach utilizes diffusion models, conditioned on class information and aided by additional Gaussian noise in the latent space, and then refines the signals through a reverse diffusion process implemented by a U-Net architecture. To validate their approach, the authors analyzed the DEAP dataset, which comprises recordings from 32 participants' brains using 32 electrodes as they responded to 40 video stimuli labeled in terms of valence, arousal, dominance, and liking. The data were first preprocessed by segmenting recordings into 1-second chunks, yielding a total of 76,800 samples. The performance of classifiers, including SVM, EEGNet, and TSception, was then evaluated on both the original and the augmented data. With CD-DPM augmentation, the synthetic data improved the TSception's valence recognition accuracy by 4.21% when the injected noise was kept minimal ($\Delta = 0.01$), while models trained with GAN- or Gaussian-only augmentation saw little or no improvement. These findings demonstrate that a signal-aware conditional diffusion approach performs better than generative baselines and highlight the importance of low-noise conditions during augmentation. The authors suggest that improving the noise scheduling and employing more sophisticated network architectures will further enhance the realism of the signals and help improve applications in EEG-based human-computer interaction and mental health care.

Alexandre and Lima (2025) [26] put forward a diffusion-based approach used to minimize the scarcity of EEG data in Brain-Computer Interface (BCI) motor imagery experiments in which the acquisition of high-quality multichannel EEG is both costly and unfeasible as it is susceptible to noise, artifact, and high inter-subject variability. Their work is specifically concerned with this weakness of the conventional generative models such as GANs which are prone to instability and mode collapse and are supposed to produce high-quality synthetic EEG channels that enable safe classification. They extract signals using the BCI Competition III Dataset V (3 healthy individuals, 32 electrodes, 512 Hz 4 classes; left hand, right hand, feet, tongue) and preprocess signals with Butterworth bandpass filter (9600

Hz, 8-30 Hz), muscle/ocular artifact detection, ICA-based correction, z-score correction and segment 1-second epochs. The presented model is a conditional DDPM where the 1D U-Net denoiser (sinusoidal time embeddings, encoder-decoder blocks, skip connections) has been trained with 1000 diffusion steps and the loss is minimized using MSE noise prediction loss. Eight channels that are less informative (Fp1, Fp2, AF3, AF4, F7, F8, T7, T8) are dropped and re-created using the nearness channels and cWGAN-GP serves as the control group. Signal-level MSE and Pearson correlation, as well as, task-level accuracy, precision, recall, and F1-score are used to evaluate under 5-fold cross-validation and paired t-tests between LR, KNN, CNN, and U-Net. DDPM tends to have lower MSE and higher correlations with cWGAN-GP and synthetic-only CNN/U-Net classification is good (95.72% and 96.31%), therefore real reconstructions, but KNN decays rapidly on synthetic data. The recommendations of future work include larger datasets, EEG-specific (spectral/phase measures) metrics, transformer integration, real-time BCI validation, and experiments of the maximum number of channels replaced without damaging performance.

The study proposes a model that involves diffusion processes and reinforcement learning for generating high-fidelity synthetic EEGs, overcoming participant burden, participant anonymity, and high collection expense (An et al., 2021) [10]. Conventional collection of EEGs can be participant-intensive and expensive, tending to yield small datasets, and can make effective development of analysis algorithms for EEGs cumbersome. To counter such a problem, in its model, a diffusion process is utilized in which both reverse and forward diffusion processes are involved. In the direction of forward, a level of randomness in terms of added noise is incorporated in actual EEGs, emulating real-life uncertainty and variance in real-life datasets. That noised information is then subjected to reverse diffusion, in which iteratively, model denoising and a model of a signal is constructed. In such a manner, both significant temporal and spectral information in the EEGs is retained, and new, synthetic datasets can be produced. In addition, a reinforcement learning agent is incorporated in the model for optimizing update strategies for model parameters during training of a diffusion model. In its role, such an agent aids in updating model parameters for a proper capture of complex dynamics in EEGs. By dynamically choosing between updates, a mechanism for reinforcement learning ensures not only accuracy in generated synthetic signals but diversity, providing a strong training base for algorithms in machine learning. Authors have utilized their model in two datasets: one, a publicly accessible BCI Competition IV 2a dataset consisting of recordings of recordings of nine subjects involved in a motor imagery activity, and a private, in-house collected experimental dataset. The combination of these datasets can generate synthetic EEG signals for a range of tasks and conditions, and generalizability can be enhanced in a model. With accuracy in classification, Kappa coefficient, and F1-score taken for evaluation, model performance showed high performance with 84.41% accuracy and 95.43% accuracy for in-house datasets. In spite of that, future work can involve diversity in datasets and computational efficiency, with future work extending model use to additional neurological conditions and computational efficiency enhancements.

Miao et al. (2024) [14] present a reinforcement learning fine-tuning strategy aimed

at boosting the diversity of images produced by diffusion models, which tend to show bias towards the training data. Although diffusion models bring many advantages, it is worth mentioning that these models usually amplify existing demographic and structural biases, resulting in poor representation in the images that are being generated. The authors apply an RL-based approach to diversity optimization, using a new type of reward function, Diversity Reward, which measures how well the generative distribution is able to cover a non-biased reference dataset. The more this reward function is optimized, the better the model will perform in terms of output diversity while maintaining the quality of the images it generates. The mentioned RL fine tuning framework is used alongside with class-conditional or text-conditional diffusion models such as Stable Diffusion which helps in mitigating the bias without negatively affecting the generative performance of the models. In particular, the Diversity Reward is computed by evaluating a collection of generated images against a reference dataset with the use of distributional discrepancy estimation methods such as MMD and mutual information. As such, it allows improving quantifiable diversity measures in a more systematic manner, with this being validated through a post sampling selection task of images where the most diverse are chosen based on the computed Diversity Reward. Experimental results yield diversity increases of 47.66% Recall improvement on ImageNet.

Wang et al. (2025) [33] have addressed the issue of learning generalizable EEG representations to a variety of tasks, especially when there is a lack of training data and the quality of the signal is low. Current models, e.g. EEGPT and LaBraM, are based on masked reconstruction goals that often do not detect rich semantic and complex patterns within EEG signals. The EEGDM self-supervised framework proposed by the authors is based on latent diffusion models and makes use of EEG signal generation as the ultimate learning goal. EEGDM uses an EEG encoder to reduce the signals and channel magnifications to succinct latent representations, which then conditions the diffusion model to generate EEG signals. This method together with channel augmentation and PCA-based latent space projection improves the signal to noise ratio and provides a solid control over the generative process. EEGDM was trained on a number of publicly available EEG datasets (PhysioMI, TSU, SEED, M3CV) and tested on downstream tasks such as abnormal detection (TUAB), event classification (TUEV) as well as motor imagery (BCIC-2A, BCIC-2B). Normal preprocessing and fine-tuning procedures were utilized, and the metrics of evaluation were the accuracy, the area under the curve, the weighted F1, and Cohen kappa. The model brought about competitive performance: EEGDM attained 80.17-percent balanced accuracy and 87.39-percent AUROC (compared to EEGPT and other baselines) on TUAB. It had a balanced accuracy of 60.81 (TUEV) and a large kappa value and on BCIC-2B, it performed as well as or better than state-of-the-art. The significance of the PCA and channel augmentation were substantiated in ablation studies with maximizing the results when both were used. Cross-dataset generalizability and high-quality EEG reconstruction were also established by the model as shown by the enhancement of correlation coefficients and the apparent separation of classes in t-SNE plots. The authors propose future research to optimize the latent space operations and pre-training to more and more diverse EEG datasets in order to enhance the generalizability and downstream performance.

Puah et al. (2025) [30] have studied how to develop powerful and computationally efficient electroencephalogram (EEG) representations of clinically relevant classification problems, including epileptic event detection. Traditional EEG foundation designs based on transformer architectures and masked token prediction often have very high computational requirements and can only achieve marginal performance improvements as they scale. In order to overcome such limitations, the authors proposed EEGDM, a new EEG representation learning model based on generative diffusion models. EEGDM uses a diffusion pretraining (SSMDP) model on a structured state-space to encode fine-grained temporal dynamics existing in EEG signals, and then uses a latent fusion transformer (LFT) to decode downstream classification. EEGDM involves a state-space model of the EEG sequences in both forward and backward time directions utilizing a bidirectional state-space model (SSM) that incorporates a trainable 128 dimensional context, either spatially in a localized fashion or noise-level, as well as embedded within the context. It is a denoising diffusion probabilistic model (DDPM) that SSMDP is trained in. Due to the dimensionality of the SSMDP latent activity outputs, multi-layer, multi-channel data is latent pooled and then condensed with a transformer based fusion module into contextual aware embeddings and processed by a transformer based classification module. The model has the advantage of law-companding to reduce the variability in EEG amplitude and uses a cosine noise schedule when training to maximize training efficiency. This paper makes it possible to define the EEGDM as an effective diffusion-based alternative to masked-token-based EEG foundation models, which provide better scalability and strength. The research directions that may be followed in the future are the minimization of the size of the model to apply them to clinical settings, the expansion of the framework to other biosignal datasets, and the use of discrete diffusion models to further develop EEG representation learning.

The authors propose a novel framework that utilizes EV-Transformer together with EG-DM to generate high-quality images from visual EEG signals based on the stimuli (Qian et al., 2023) [16]. To evaluate the effectiveness of their proposed method, the authors employ two primary metrics: Classification Accuracy and Inception Score. The evaluation metric Classification Accuracy shows how well the EV-Transformer decodes EEG signals while performing its decoding activities. The model's success in visual features identification from EEG-derived data emerges from this evaluation metric. The successful ability of the EV-Transformer to extract meaningful EEG signal information becomes evident through its greater classification accuracy. The framework exhibited remarkable performance by reaching 99.80% average classification accuracy for forty-class data and 92.07% for four-class data. EEG signal decoding by the proposed EV-Transformer highlights its effectiveness in direct brain-mapping applications for brain-computer interface systems. The Inception Score serves as the metric to assess both clarity and diversity within images produced from EEG-Guided Diffusion Model operations. The Inception Score combines two metrics to assess both the realistic appearance of generated images while capturing their wide variety of outcomes. The model achieves clarity by matching electronic signals with recognized visual inputs yet displays diversity through its extensive range of visual outputs. Image quality improved according to elevated Inception Scores because these scores measure sharp visual outputs with many distinctive image variants. The NeuroDM framework achieved an Inception Score of

15.04 on forty-class data and 8.67 on four-class data which indicates superior quality image generation skills. The current research presents ongoing challenges regarding performance stability between users and image variety generation that future studies may address.

Yang et al. (2024) [20] examined the longstanding problem of the reconstruction of high-fidelity visual stimuli with non-invasive EEG signals, which are characterized by low spatial resolution and intrinsic noise of EEG responses, and by the relative lack of realism and semantic consistency of the previous image-generation algorithms. In order to overcome these shortcomings, the researchers proposed the EEGConDiffusion framework, which combines powerful EEG features extraction with a conditional diffusion model to enable converting EEGs to images directly. The method focuses on a new convolutional neural network, EEGConvNet, which has individual temporal and spatial convolutional layers and numerous residual blocks to thoroughly capture discriminative EEG features to reduce over-fitting and gradient-related problems. These features are modeled and embedded with a positional-encoding network and a pre-trained FrozenCLIPEmbedder that have been trained so as to align them with the input expectations of a fine-tuned Stable Diffusion (SD) model built on latent diffusion and a U-Net architecture. The framework was tested on a public EEG image classification dataset, 12,000 EEG trials of 6 subjects individually viewed 2,000 images of 40 ImageNet classes, and in total, 128 channels of EEG were recorded. The preprocessing procedure involved 1-70 Hz and 5-95 Hz filtering, with the latter being used in the later experiments as the optimal range of features to extract. The measures of performance were classification accuracy, kappa score, 50-way top-k accuracy, and Inception Score (IS). EEGConvNet attained a mean classification rate of 67.97 percent, outshining LSTM, ChannelNet, and EEGNet by 56.34, 47.79 and 30.60 percent respectively. Statistically significant increases in the kappa values and Wilcoxon signed-rank tests (p ¡ 0.05) were also shown by the model. In image generation, the framework generated high-quality images with significant semantic consistency, which is demonstrated by better Inception Scores and top-k accuracies on the generative models, as well as, compared to baseline generative models. The paper emphasizes the need of multimodal large models to decode the EEG signals and produce realistic images, and the EEG-ConDiffusion framework displays high generalizability and effective feature extraction. Limitations: The study requires better EEG-image feature matching and higher inter-subject transferability. Optimizing multi-level feature alignment and transfer learning are the ways more research could be enhanced to improve EEG-to-image decoding.

Chen et al. (2024) [11] examined the enduring drawbacks of the limited availability of EEG data and lower quality of the synthetic EEG signals used to augment the data as it is utilized in classification tasks. Previously generated methods, such as GANs and diffusion models, generally produced low quality images and failed to consistently enhance classification, especially when the data generated was added directly. In order to overcome these limitations, the authors proposed a Transformer-based denoising diffusion probabilistic model (DDPM) specifically optimized to synthesize EEG signals, combining Multi-Scale Convolution (MSC) and Dynamic Fourier Spectrum Information (DFSI) modules to better represent the variability of EEG frequency content and to increase the stability of training. It used a constant-factor

scaling technique in preprocessing to retain the amplitude information and ensure that the models converged. The main augmentation methodology entailed the random recombination of portions of the generated and original EEG recordings in the time domain to generate vicinal samples thus decreasing empirical and vicinal risk in training. Smoothing of the synthesized data was done to reduce the chances of spreading the wrong labels. It was tested on four EEG datasets available publicly (Bonn (epilepsy detection), SleepEDF-20 (sleep stage classification), FACED (emotion recognition) and Shu (motor imagery)) with EEGNet and EEG-Conformer as classification backbones. As a downstream measure, the Fréchet Inception Distance (FID) was used to measure the quality of performance of the signals and the accuracy of classification. The EEG Diffusion Transformer recorded the lowest FID scores in all four datasets, which is an indicator of high-quality signal compared to GAN, VAE, and UNet-Diffusion baselines. The accuracy improvement with the augmentation technique was highly significant: 14.00 percent on Bonn, 6.38 percent on SleepEDF-20, 9.42 percent on FACED and 2.5 percent on Shu. The effectiveness of MSC and DFSI modules, label reconstruction and the composite loss function were supported by ablation studies. In general, the approach exhibited strong generalizability with a wide range of tasks and network architectures, and the future studies will be focused on further validation to other time-series areas.

Chaudhary and Jaswal (2021) [3] performed an extensive survey of EEG-based emotion recognition with the focus on the DEAP dataset and emphasized the need to make the classification of emotions accurate to achieve the successful interaction between humans and machines. Traditional facial expression techniques often confuse affective states due to their similarity, e.g. sadness, frustration, or anger and EEG offers a clearer and more objective interpretation of emotion because of the similarity between the expression and the neural activity. DEAP, a widely-used emotion analysis data set, consists of EEG data of 32 individuals watching 40 music videos, accompanied by subjective ratings of valence, arousal, dominance, liking, and familiarity. The authors critically reviewed various existing methods that use DEAP, such as feature extraction pipelines with classifiers such as support vector machine (SVM), bidirectional long short-term memory networks (BiLSTM), long short-term memory networks (LSTM), convolutional neural networks (CNN), and eXtreme Gradient Boosting (XGBoost). The empirical results indicated uneven performance: CNN-based models exhibited better performance with about 83-85 percent per valence and arousal classification; meanwhile, LSTM and BiLSTM architecture, which represents strong performance in temporal modelling, were found to perform at the middle of the 60s to middle of the 80s. The quantification of the dynamic properties of EEG signals by feature-based methods, which focused on power spectral density and entropy measures, widely prevailed. The comparative study on various studies showed that the combination of various frequency bands with advanced classifiers improves the effectiveness of emotion recognition. Some of the issues that the authors also discussed included inter-subject variability, the need to have strident data preprocesses, and the inherent complexity of the interpretation of the EEG signal. To further the discipline, the authors suggested an artificial neural network (ANN) classification scheme based on salient time domain features such as correlation, mean, variance, kurtosis and skewness extracted on separate frequency sub bands. Down-sampling together with band-pass filtering was used

in preprocessing sequences to retain key EEG rhythms. Empirical findings have shown that ANNs are capable of achieving high levels of both valence and arousal categorisation and at the same time maintain model simplicity and computational efficiency. Comprehensively, this tiring literature review highlights the importance of EEG modalities in emotion recognition and pre-empts the need to conduct future investigation to examine hybrid deep learning structures and optimised feature extraction schemes, thereby attaining increased accuracy and strength on the DEAP dataset.

Mouazen et al. (2025) [29] discovered the affective states detection in electroencephalographic (EEG) and used hybrid machine-learning models which were trained on the DEAP data. The first thing that they wanted is the categorisation of the valence and arousal dimensions which are known to be the fundamental elements of emotional experience. To overcome the existing challenges associated with the classification accuracy, interpretability of models, and practicality of deployment in real-time, the authors combined the classical machine-learning algorithms with deep-learning models. Preprocessing included band-pass filtering, artifact reduction and temporal partitioning and was followed by feature extraction, with special emphasis on frequency-domain features especially power-spectral density (PSD) at six canonical bands spanning Delta to Gamma. These type descriptors have been comprehensively reported as neurophysiologically informative indices of affective state and the use of standardisation methods allowed inter-subject comparability. K-Nearest Neighbours, Support Vector Machines, Decision Trees, and Random Forests formed the basis of the baseline models but they had weak generalisation due to overfitting. To overcome the problems of high dimensionality and complex temporal interdependencies inherent in EEG data, the authors developed a hybrid deep-learning pipeline, which combines autoencoders to dimensionality reduce it, transformer encoders to capture long-range dependencies, and bidirectional Long Short-Time Memory (LSTM) layers to capture short-term dynamics in the temporal dynamics. The highest accuracy was achieved with a Transformer-LSTM additive fusion architecture, with an accuracy of 73.58 on the test-set of 73.58. Interestingly, recurrent neural networks, especially the bidirectional LSTMs achieved the highest possible accuracy of up to 94 percent beating the Convolutional Neural Networks and Gated Recurrent Units, thus validating the idea of using recurrent neural networks to model sequential EEG data. To further enhance the interpretability of the models, the authors used SHapley Additive explanations (SHAP), which revealed the Theta and Alpha frequency bands as the most effective factors influencing the prediction of the affective state, which supports extensive neuroscientific findings. The findings, therefore, highlight the feasibility of hybrid deep-learning models to perform real-time emotion recognition at provably high accuracy whilst being interpretable. Future directions of the research would be to scale the data to enhance generalisation, use graph-based neural networks to learn the inter-electrode spatial relationships and combine multimodal physiological signals to enhance predictive accuracy further. The authors promote personalised modelling and real-time deployment as the key steps to the practical use of EEG-based emotion detection systems.

Jiang et al. (2024) [12] address the challenge of predicting future trends in multi-

channel EEG for early prediction of epileptic seizures, a function hardly addressed in present algorithms for diagnostics. Traditional models care less about predicting future trends, and only contribute towards diagnosing and classifying present-day EEGs, but not predicting future trends, which is important for early intervention. In a new work, in a new contribution, an EEG-DIF model addresses such a challenge through reformulation of forecasting into an image completion problem, with spatio-temporal correlations and future trends in EEGs represented and learned effectively. With a use of a Denoising Diffusion Implicit Model (DDIM) with a U-Net backbone, with a mechanism for completing noised images and predicting future trends in a signal, EEG-DIF model addresses multi-channel EEG complexity through a mechanism for completing noised images and predicting future trends in a signal. Model predictive accuracy and model explanatory power, with an objective of minimizing errors and maximising explanatory power, in terms of a use of Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and a use of a Coefficient of Determination (R^2) evaluation metric, is considered for testing with Siena Scalp EEG Database, with its recordings for 14 subjects (9 males, 5 females) in disparate age groups, its data including several electrodes for EEG and one or two for EKG, in addition to patient information including a classification with/from epilepsy, a number of seizures, a number of channels for EEG, and a duration for a record of approximately 128 hours with 47 seizures. As per work, an astounding accuracy in early prediction of 89% for predicting seizures is achieved with a use of the EEG-DIF model. However, the performance of all channels is not even, with individual channels having relatively poor performance in evaluation, and even a potential relation with channel ordering in an image representation. In future, future work is proposed concerning channel ordering and its consequence for prediction performance and extending the function of the model with further channels and optimized inference for application in a medical environment. In this work, the utility in utilizing early stage epilepsy seizure forecasting with generative diffusion models is proven, with avenues for future improvement.

Wang et al. (2024) [18] suggested the unsupervised EEG-based seizure anomaly detector, SAnoDDPM, which is not limited by the size of expert-labeled data, unlike many predictors of the occurrence of EEG-based seizures, thus it is difficult to obtain, time-consuming, and easily subject to subjective labeling bias and overfitting during limited-data conditions. The authors consider the problem of seizure detection more of an anomaly detection problem, where the model is only trained using EEG free of seizure, and then detects the seizures as a deviation of normal patterns. By the first time using Denoising Diffusion Probabilistic Models (DDPM) to detect seizures they introduce a new approach to the model using DDPM that is trained on normal latent images and then uses the model to restore the use of seizure-affected EEG by restoring the seizure segments into resembling normal-looking latent images, which detects the presence of the seizure by comparing pre- and post-restoration EEG. The process transforms EEG into spectrograms 2-dimensional (STFT) then compresses them with VQ-VAE into 2-dimensional latent codes and uses a noise predictor with U-Net-based architecture and sinusoidal time embeddings (T=1000 diffusion steps). The most important novelty lies in the fact that variable lower bounds on Markov chains and KL divergence are used to create anomaly masks (optimal at steps 400-600), which steer the recovery process. On CHB-MIT (23 pediatric patients, 256

Hz, approximate 958 h), and TUH (20 selected patients out of 6724 patients 30 min segments) experiments, both had high performance; on CHB-MIT, 98.77 percent accuracy and 93.91 percent event sensitivity (0.79 FDR/h); on TUH, 98.16 percent and 92.11 percent accuracy and event sensitivity (0.93 FDR/h) and low. Future directions consist of seizure prediction and generic cross-patient model rather than the patient-specific training.

Shao et al. (2025) [31] explored the long-term challenge of effective electroencephalography (EEG) artifact elimination, with special attention to the situation with high correlations between artifact and EEG signals and the use of single-channel information. Generalisation or maintenance of relevant neural activity in such cases is often incapable of conventional regression algorithms, blind source separation algorithms, or joint algorithms. Hence the authors came up with D4PM, dual-branch driven denoising diffusion probabilistic model, which integrates multiple-type artifact removal by jointly modelling clean EEG and artifact distributions. D4PM consists of two trained separately conditional diffusion models, one focused on clean EEG, and the other on artifacts. Both models use a dual-path structure that is driven by joint Dual-FiLM layers to allow the capture of both motions and structures. The framework uses a joint posterior strategy whereby both branches combine complementary priors but with data-consistency constraints, thus making it easy to collaboratively and high-fidelity replicate EEG signals. The EEG records, OG records, EMG records, and ECG records in the mixed datasets acquired on EEGNet and MIT-BIH Arrhythmia repositories were used to conduct the experimental assessment, which consisted of 4,514 EEG records, 3,400 EOG records, 5,598 EMG records, and 3,600 ECG records respectively. These data were divided into training (80 per cent), validation (10 per cent) and test (10 per cent) subsets. To evaluate the quantitative assessment, it was assessed in the time domain (relative root-mean-square error RRMSEt) and spectral domain (relative root-mean-square error RRMSEs) and correlation coefficient (CC) and the signal-to-noise ratio SNR. The performance of the D4PM model was state-of-the-art, especially on the elimination of EOG artefacts (CC=0.995, RRMSEt=0.098, SNR=20.983). It out-performed all the baseline approaches, such as FCNN, SimpleCNN, EEGIFNet, GCTNet, and EEGDfus. The joint anterior sampling mechanism and the dual-branch architecture were both found to be effective in the ablation studies. The modelling of artifacts also helped in achieving better performance especially in processing EMGs and ECGs. The single generative strategy used in D4PM facilitates effective artifact elimination in single-channel and mixed-artifact conditions as well, and the elimination has excellent generalisation to unknown and cross-domain samples. The authors envision the lines of future research in mixed-artifact modelling and more extensive application in real-world EEG analysis applications.

Yuan et al. (2024) [21] proposed a new 3D diffusion model, ReMiND, to mitigate the issue with unavailability of information in structural MRI in longitudinal Alzheimer's disease (AD) studies. Missing information in imaging, sometimes a result of participant dropout, motion, or technological failure, can make useful markers challenging for studying progression in AD challenging. ReMiND utilizes denoising diffusion probabilistic models (DDPMs) to generate missing whole-brain images with one, several, or any observable MRIs of an individual, accurately depicting temporal and

spatial trends in imaging. ReMiND was validated using T1-weighted MRI information extracted from Alzheimer's Disease Neuroimaging Initiative (ADNI) databases, a collection including demographics, clinic, and imaging information about participants with normal cognition, mild cognitive impairment, and AD. Bias correction and preprocessing through registration preserved uniformity and trained with concatenated tensors of past, current, and future data points in an attempt to reduce the loss during generation. ReMiND performance compared with both with and with a general varitional autoencoder model considered SSIM and PSNR factors, with performance reflective of high fidelity in restored images, with SSIM values of roughly 0.974 for ReMiND-P and 0.984 for ReMiND-PF and PSNR of 31.786 dB for specific settings, depicting high semblance with real-life images and accuracy in atrophy progression estimation in and over time. Volumetric change prediction in volumetric regions with anatomic impact displayed a clear standout, overcoming a significant weakness in current imputation approaches. ReMiND is a breakthrough in neuroimaging but with its weaknesses including nonequivalence in training when including skult-stripped structures and in use in cases with complex cases with missing, possibly with an impact for extrapolation performance in real cases with complex missing scenarios and settings not included in model training. Future research will enable refinement of preprocessing algorithms, generalize the model to include variable windows in time, and assess its accuracy with high resolution datasets. Integration with clinic-imaging and multi-modal datasets can make predictive accuracy even better, and allow for sounder modeling of disease progression in AD. By completing gaps in image progression, ReMiND not only brings a strong alternative to dealing with missing values but opens a platform for creating complex predictive and biomarker models in studies of AD.

2.3 Summary of Key Findings

Missing or corrupted EEG data (which results from electrode detachment, physiological artifacts as well as intentional rejection of EEG segments) fatally compromises the reliability of neurophysiological analyses, especially for applications that require temporal continuity (e.g. emotion recognition, clinical diagnostics, brain-computer interfaces). Addressing this challenge, recent advances in generative deep learning have agreed on Denoising Diffusion Probabilistic Models (DDPMs) for a stable and high-fidelity counterpart of Generative adversarial networks (GANs). diffusion-based frameworks have effectively shown to model the complex spatiotemporal dynamics inherent to a multi-channel brain signals shown superior performance. Contemporary approaches increasingly combine DDPMs with complementary architectural components: Variational Autoencoders (VAEs) for controlling the compression of the latent space for efficient computation during synthesis as suffered under data imposing a demand for conversions Experts such as Graph Neural Networks (and more particularly Graph Attention Networks called GATs for explicit modeling of inter-channel functional connectivity as learned spatial relationships Transformer architectures for temporal dependencies between long-range signal sequences. Beyond generation under the very limited conditions, masked autoencoder paradigms such as (MAE-MI, EEG2Rep) use self-supervised generating pretraining on unlabeled data and strategic masking and reconstruction objectives to learn

generalizable EEG representational relationships and conditional diffusion models allow for subject-specific, session-specific, and class-conditioned synthesis that are critical in addressing issues of data imbalance and inter-subject variability. Notably, hybrid approaches using both the preprocessing in the spectral-domain (e.g. Short-Time Fourier Transform, power spectral density features) and diffusion models maintain frequency-band characteristics for neurophysiological validity, and the combo chain of reinforcement learning optimizes generative processes for domain-specific objectives e.g. class accuracy or clinic favorability. Evaluation frameworks have now encompassed both signal-level fidelity metrics (Fréchet Inception Distance, multi-scale structural similarity, power spectral density distance) and downstream task performance by verifying the fidelity of the amount of high-quality synthetic EEG augmentation which consistently boosts the generalization of models, especially with low-data regimes, while preserving spectral features between canonical frequency bands of the brain.

Paradigm shift from rejection of artifacts along with data loss to strong signal reconstruction which is allowed by these methodological advances together, also these methodological advances enables personalized neurodiagnostics, data privacy preserving synthetic data sharing in clinical research and realtime BCI applications.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

The final system is based on a 1D U-Net-based DDPM model enhanced with Singular Value Decomposition (SVD) and Graph Attention Networks (GAT) for improved spatial dependency capture, was trained using a DEAP EEG data set.32-channel EEG signals are processed, and extreme values are clipped and the signals are standardized. To train, a CUDA-capable GPU, minimum 8GB+ VRAM, and Python libraries such as PyTorch and NumPy are needed. The model is a modular one with about 6.3M to 7.4M parameters based on configuration. Preprocessing and model training scripts are reusable and well-documented.

3.2 Societal Impact

The project can make a lot of difference in healthcare, especially in mental health tracking and brain-computer interface. EEG reconstruction can facilitate diagnosis as well as treatment of neurological diseases. Nevertheless, abuse of EEG information may give rise to issues of ethics and privacy. Fair access to technology plays a significant role in avoiding digital health disparities. Regulations and moral rules must be made to be used responsibly.

3.3 Environmental Impact

The association of this project with the environment can be mainly linked with the digital processing of EEG datasets. Training the model requires resources of the GPU, which consumes electricity and adds to the carbon footprint. Given the complexity of the model, made up of the combination of DDPM, GAT, and SVD-based decomposition modules, it has taken well over 6 hours to train on GPU, indicative of its ability to process multi-channel EEG data and generate high-fidelity reconstructions. While this training time is significant, it is relatively efficient relative to vast-scale models that are used in the processing of images or language. The computational requirements of the proposed model are supported by its ability to reconstruct not only time but also space information using EEG signals,as well as

leveraging low-rank structural direction using SVD to enhance the quality of reconstruction without necessarily having to use model sizes that are too large. By making use of energy efficient hardware and by utilizing common compute resources, the impact on the environment is minimized. No physical or hazardous materials are involved, for a low environmental footprint and EEG signal analysis advancement.

3.4 Ethical Issues

EEG data is confidential and sensitive hence it must be handled accordingly in terms of privacy and confidentiality. The evaluation of the bias of models across populations must be undertaken. The application of artificial or replicated brain signals is questionable of control and manipulation . Transparency and interpretability should be embraced by the developers, especially in the health sector. Both should be under ethical supervision phases of research and deployment.

3.5 Standards

The project has complied with the laid down requirements with regard to data processing and model development. The use of EEG data is in line with GDPR in data privacy and consent. Neural data processing obeys standard pipelines of BCI signal preprocessing. Software development is written in Pythonic style, modular design, and uses Git version control. IEEE long-term deployment guidelines of neural interface systems are contemplated health-related tools.

3.6 Project Management Plan

The project was organized in a milestone-based plan for months 5–8. During Month 5, the focus was on collecting and preparing the dataset, including cleaning and transformation. Month 6 was dedicated to designing and implementing the initial model based on Denoising Diffusion Probabilistic Models (DDPM). In Month 7, the first version of the model was finalized, and a Pre Thesis 2 Report and Presentation were prepared to showcase progress and preliminary results. Month 8 was used to present the Pre Thesis 2 and start evaluating the model against baseline criteria.
From Month 9–12, thesis defense preparation took over. During this phase, considerable improvements were implemented to enhance the model and its overall performance, including handling missing data and enhancing its overall functionality. The most important improvements consisted of adding Graph Attention Networks (GAT) to learn how to capture inter-channel spatial dependencies and functional connectivity patterns in the 32-channel EEG modified signals and supplemented with a Singular Value Decomposition (SVD) to learn low-rank spatiotemporal components. The development of multi-head graph attention layers (4 heads, 64 hidden dimensions) on each encoder level and the SVD preprocessing module (decomposing EEG signals into spatial (U matrix), temporal ($V^\top$ matrix), and importance weights (singular values, top 16 components kept of the original signal features), and their fusion with the original signal features implemented through adaptive learned combination mechanisms resulted in this change of the baseline DDPM to the optimized DDPM+GAT+SVD architecture.

In Month 9, the model was evaluated and compared with baseline results as well as other existing methods. On the basis of such comparisons, optimisation at Month 10 was carried out to further optimize the accuracy and reliability of the developed DDPM+GAT+SVD architecture, such as fine tuning the fusion parameters, refining the number of SVD components and fine tuning the fusion of SVD features with GAT. Month 11 saw the thesis being finalized incorporating the improvements and preparing towards the submission. The project was completed in Month 12 with the completion of the thesis defence and submission.
The team relied on a Kanban-style to-do list to track progress and used Git for version control. The main resources required were computing power for training and free, open-source tools. There were no additional expenses aside from GPU usage.

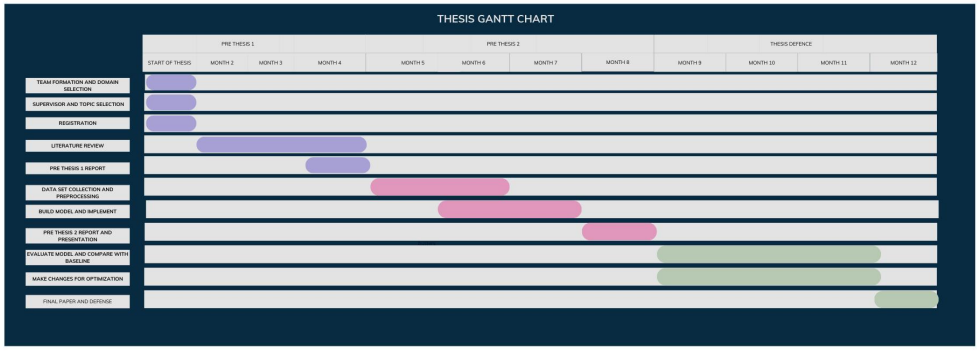


Figure 3.1: Thesis Work plan Gantt Chart

3.7 Risk Management

The major risks are training instability, overfitting, or inability to generalize across subjects. To reduce this, normalization plans, EMA regularization, and robust techniques of evaluation were used. The other risk is the data leakage or ethical misuse, this is covered by anonymizing inputs and the consent procedures. Hardware reproducibility testing helped to reduce failure or CUDA incompatibilities on several systems. Model failures can be caught early with the use of visualization.

3.8 Economic Analysis

Cost-effectiveness of DDPM+GAT+SVD model is outstanding with low economic barriers for implementation. The entire development used freely available resources - the open-source DEAP dataset came with no licensing fees and the training was done on Kaggle's free P100 GPU allocation - so there was no need for hardware cost at all. Model training took about 100 minutes (50 epochs at about 5 min-

utes/epoch), with no monetary cost, while the computational cost of inference is negligible (˜3 seconds for reconstruction of 640 samples of EEG) Compared to manual artifact removal by trained clinicians (estimated $5–10 per patient at 15 minutes * $20–40/hour labor cost), our automated approach has 80.6% better accuracy (RMSE 0.110 vs 0.570 baseline) at a cost of effectively zero operation cost once deployed. The model's 19.5M parameters require just 78MB storage, making it possible to implement the model in standard clinical workstations - it does not need dedicated IT infrastructures. The main limitation in terms of economical aspects is its once off need for GPU access for the initial training, but the availability of free platforms to run cloud systems (Kaggle, Google Colab) and the capability to run the model on consumer-grade GPUs ($200–500) make it economically viable for resource-constrained research institutions and clinical settings in developing regions.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

The chapter has a stepwise description of the process by which a diffusion-based generative model in the reconstruction of electroencephalographic (EEG) signals was constructed and tested. It discusses preprocessing of data, precomputed Singular Value Decomposition (SVD) for feature extraction, theoretical context of the Denoising Diffusion Probabilistic Model (DDPM), the theoretical model structure of 1D U-Net, training, and evaluation. Baseline correction and splitting into overlapping parts of 640 samples are pre-processed in the DEAP data set and normalization per channel. As part of loading the dataset, SVD is precomputed to split the normalized signals into spatial (U matrix), singular value (Σ), and temporal (V^T matrix) components, truncation of which keeps 16 out of 32 components in order to reduce computation time and noise. DDPM uses steps of diffusion (300 diffusion timesteps iteratively) to generate and predict noise (ε) by using a forward diffusion process and a reverse denoising process, whereas the 1D U-Net with SVD-derived features and Graph Attention Networks (GAT) considers both the time-dependent and place-dependent features of the EEG data. The pre-computed SVD features (U and V^T matrices) are initially processed using an SVDFeatureExtractor module, which are then combined with the raw signal features and finally operated on using encoder blocks that consist of spatial inter-channel attention GAT layers and temporal convolutions. The model uses a masked reconstruction strategy in evaluation, which can predict 30–40% of central missing parts of the signal. This model is trained in 50 epochs with Adam optimizer and MSE loss and tested with such metrics as MSE, RMSE, MAE, PCC, and PSD. Python was used in all processes and was tested on the source code.

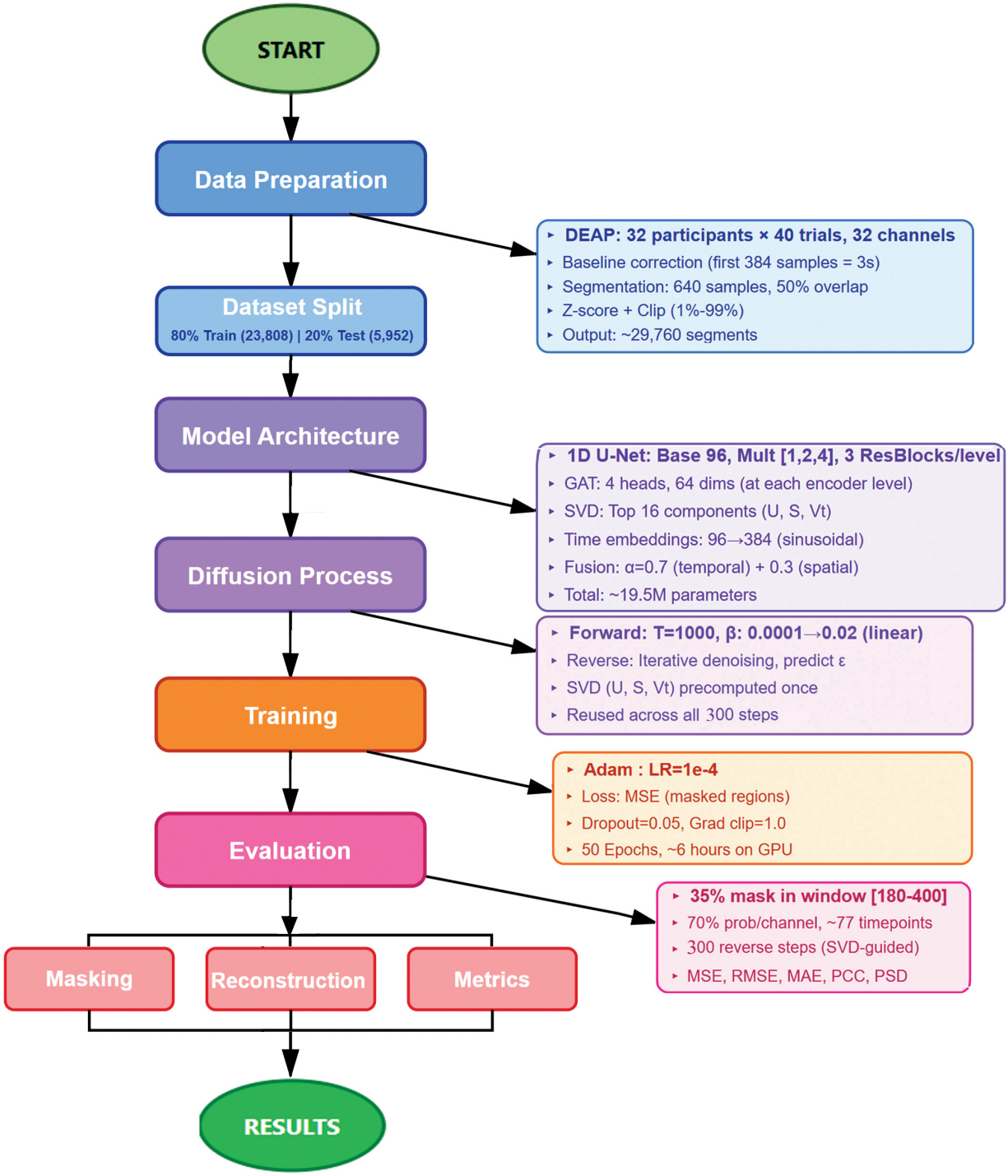


Figure 4.1: Methodology Flow for EEG Generation Model

4.1.1 Dataset and Preprocessing

The paper used the DEAP (Database of Emotion Analysis with Physiological Signals) data that consist of EEG data of an exposure of 32 individuals to 40 different emotional video stimuli. The respondents watched 40 music videos of one-minute length, which were set to evoke certain emotional reactions during the experiment, and were monitored with EEG signals during the entire time of watching the videos.

Out of the 40 initial channels recorded (applying the 10-20 electrode system established internationally), the 32 EEG channels remained (the rest of 8 peripheral physiological signals (EMG, GSR, respiration, plethysmograph, and temperature) were dropped. This gave data matrices of 32 channels by 8064 timesteps / trial, which translates into a minimum record duration of 63 seconds at 128 Hz.
The first preprocess method was that of baseline correction in order to eliminate everyone's psychological changes as well as DC offset artefacts. The mean amplitude in the first 384 timesteps (equivalent to a 3-second pre-stimulus baseline period) was then calculated, separately on averaging each of the 32 channels and this baseline was then subtracted off the full corresponding channel record to facilitate zero-baseline consistency across all the trials and subjects.
To make the amount of the training data larger and introduce a broad range of the possible temporal patterns that emotional responses to the EEG can have, the continuous 63-second EEG records were divided into overlapping windows. In particular, a sliding window processing (overlap = 50 percent) was used to divide each trial into 640 timestep(step = segment_length // 2 = 320 timesteps) segments (which is 5 seconds at 128 Hz) in order to divide the trial into smaller segments. This type of segmentation keeps the temporal continuities in the segments and affords the model adequate context to learn short term and long term temporal relationships in the dynamics of the EEG signal.
Following segmentation, channel-wise z-score normalization was used for the standardization of the signal amplitude distributions of different channels and participants. The global mean and the standard deviation were calculated separately for each of the 32 channels over the whole training data set, and all signal values were normalized to have zero mean and unit variance. This normalization step ensures that the training of the model is not biased towards those channels that have inherently larger amplitudes (occipital channels for one), and that the gradient flow is stable when the model is being backpropagated. To further account for the effect of extreme outliers and measurements artifacts (such as a spike of electrode impedance or a motion artifact) the normalized distributions were clipped at the 1st and 99th percentile. This clipping operation ensures that all the preprocessed signals are of a statistically consistent range (of about ± 2.58 standard deviations), which is extremely important for training stable deep learning models and ensuring that gradient explosion is avoided during the early stages of training.
The preprocessed data set was then divided into training and testing data sets, 80% and 20% respectively, using the random sampling method without replacement (train_size = int(0.8 x len(dataset))), by keeping the statistical independence between the different splits to achieve unbiased data evaluation and avoid data leakage in the process of testing data.

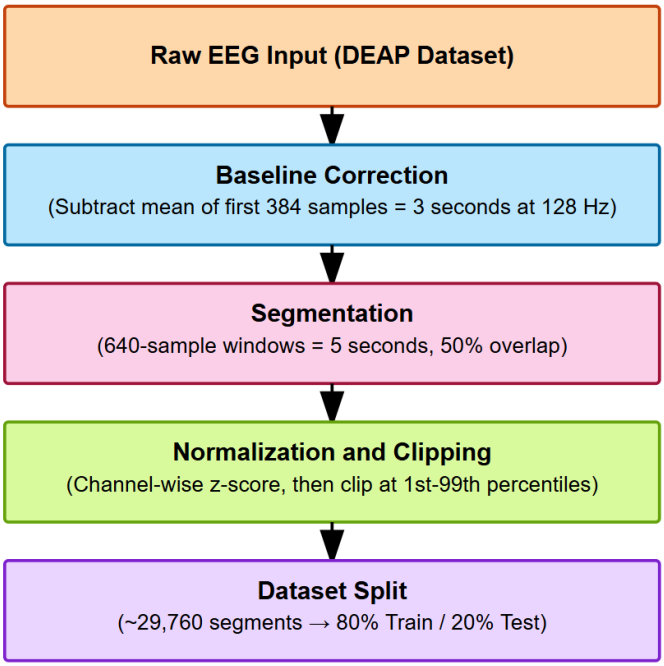


Figure 4.2: Data Preprocessing Steps

4.1.2 Denoising Diffusion Probabilistic Model Framework

The generative reconstruction model is built on the basis of the Denoising Diffusion Probabilistic Model (DDPM), which comprises forward noising model and learned reverse denoising model working in discrete timesteps. The forward diffusion conditionally subjects the image to Gaussian noise over 300 discrete timesteps, where the nerve of the linear variance schedule generated by beta values rises (β_t) in steps ranging between 0.0001 and 0.02. Such gradual contamination of the clean EEG signal leads to an ultimate timestep at which point the signal is of an approximation pure Gaussian noise. The noisy signal at timestep t is calculated by:

$$x_t = \sqrt{\bar{\alpha}_t} \times x_0 + \sqrt{1 - \bar{\alpha}_t} \times \epsilon \tag{4.1}$$

where $\bar{\alpha}_t$ is the cumulative product of alpha values (with $\alpha_t = 1 - \beta_t$) and ϵ represents random Gaussian noise.

By making this formulation, it is possible to sample directly at any timestep, without taking the noise addition process sequentially. A neural network ϵ_θ Diffusion step reveals a noise term, so the reverse denoising process is controlled by a separate neural network. During training, the model minimizes the Mean Squared Error (MSE) between the actual added noise (ϵ) and the predicted noise (ϵ_θ), thereby

learning the reverse trajectory from noisy observations back to clean signals. To achieve training stability, gradient clipping which has the maximum norm of 1.0, is used during backpropagation and the Adam optimizer with the learning rate of 1×10^{-4} and the cosine annealing schedule is used to ensure smooth convergence in the 50 epochs allowing the model to learn the reverse diffusion dynamics that know how to reconstruct EEG signal.

4.1.3 Hybrid Architecture: 1D U-Net with Graph Attention Networks and SVD Integration

The developed model is a hybrid architecture consisting of a 1D U-Net backbone integrated with precomputed Singular Value Decomposition (SVD) features and Graph Attention Network (GAT) layers that can be used to capture both temporal dynamics and spatial inter-channel relationships in EEG signals simultaneously. Prior to the encoder-decoder processing, two parallel input streams are passed to the model: first is the raw normalized 32 channel EEG signal, and the second is precomputed SVD components ($U \in R^{32\times16}$, $S \in R^{16}$, $V^\top \in R^{16\times640}$) that decompose the EEG signal into orthogonal spatial-temporal bases. They are processed by an SVDFeatureExtractor module which develops these components using three specialized paths: spatial patterns (U matrix) are processed using 1D convolution ($n_{\text{components}} = 16 \rightarrow \text{out\_channels}/2 = 48$) with group normalization (4 groups) and SiLU activation function followed by upsampling to match the temporal dimension; temporal patterns ($V^\top$ matrix) are processed using 1D convolution (kernel size 3, padding = 1) with the same normalization; singular values (S vector) are transformed using a 2-layer MLP ($16 \rightarrow 96 \rightarrow 96$) in order to generate importance. The processed triplet SVD features (96 channels), then, are concatenated with the raw signal features (96 channels) from the first convolutional layer, and are then fused using a 1×1 convolution in the presence of group normalization (8 groups) and SiLU activation to get a fusion feature of 96 channels based on both the data-driven (raw signal) and the mathematically-derived (SVD) feature spaces.

The U-Net is also an encoder-decoder architecture that has a symmetric skip connection between its two counterparts, and the symmetric skip connection is used to make sure that multi-scale features are retained during the reconstruction process. This encoder pathway comprises three levels of hierarchy, with methods of channel multiplication of $[1, 2, 4]$, increasing from 96 to 384 channels at the bottleneck. Spatial resolution is reduced by a half at every level with the help of strided 1D convolutions (kernel size 3, stride 2).

In every encoder level, there are special encoder blocks that combine both temporal and spatial processing. In particular, the extractor of local temporal patterns is a series of residual blocks (three in this configuration), equipped with grouped normalization (8 groups), and SiLU activation functions to extract local temporal patterns, while multi-head graph attention layers (4 attention heads, 64 hidden dimensions) operate on the postfusion latent features (not directly on SVD-U matrices), with EEG channels as nodes in the graph with learned weights as attention that capture functional dependencies between brain regions through a fully-connected attention mechanism without explicit adjacency matrices. Sinusoidal positional encodings of the diffusion timestep are first generated in 96 dimensions using sine and cosine functions, then transformed to a 384-dimensional embedding (model_channels $\times$ 4)

through a two-layer MLP with SiLU activation to generate time-conditional encodings that are injected into each residual block to condition the model on the current diffusion state. The decoder architecture is similar to the encoder but upsampling is performed by transposed convolutions with the kernel size of 4 and a stride of 2, and the skip connection is concatenated with skip connections on the same level at the identical encoder across different fine-grained signal detail scales to an equal level of the signal. A sigmoid-activated weighting ($\alpha \approx 0.7$) learnable fusion mechanism with the combination of the temporal features of residual blocks and spatial features of GAT layer at each encoder stage allows the model to dynamically trade-off between local temporal and global spatial associations.
To avoid overfitting, dropout regularization (5%) is used throughout the architecture. The last output-layer consists of grouped normalization, SiLU activation, as well as 1D convolution that maps the 96-dimensional features to the original 32-channel EEG signal space, which finalizes the end-to-end reconstruction process.
The combination of features extracted using SVD allows the model to adopt combined representations from learning (the fusion of two representations, achieved by U-Net+GAT) and mathematical priors (SVD decomposition providing explicit spatial-temporal structure), where the components of the SVD guide the diffusion model to physiologically plausible reconstructions while the GAT layers allow to guide these structures towards segmentations in an adaptive manner, according to the learned inter-channel dependencies in the fused latent space.

4.1.4 Model Training and Implementation

Training was performed for 50 epochs (batch size of 16 samples) with the Adam optimizer (learning rate: 1×10^{-4}) having the capacity to adapt to changes in parameters. During the training, a cosine annealing learning rate schedule was used, and the learning rate was gradually decreased starting with the original initial learning rate to almost 0 at the last epoch, allowing the convergence to be smooth without a separate warm-up phase. At the data loading stage, for each training batch there are 4 parts that include (1) clean EEG segments (32×640), (2) precomputed U matrices (32×16), (3) precomputed S vectors (16), and (4) precomputed V^T matrices (16×640), which are loaded simultaneously from the preprocessed data set to avoid extra time-consuming SVD computations in the training stage. These precomputed SVD components of precomputed components are frozen (non-trainable), provided as fixed structural priors and the SVDFeatureExtractor drawing them is trainable.
In every training step, the dataloader provided clean EEG segments along with their corresponding precomputed SVD components to the model and randomly chose diffusion timesteps (uniformly sampled between 0 and 299) per sample in the batch to allow the network to learn the prediction of noise at all stages of the diffusion process. The clean signals were corrupted with Gaussian noise generated based on the forward diffusion equation and the U-Net-GAT-SVD architecture came up with an estimated added noise conditioned on the noisy input, timestep embedding and the precomputed SVD components (U, S, V^T) which provide additional spatial-temporal structural information to guide the denoising process.
The objective of the training was the Mean Squared Error (MSE) between the predicted noise and the actual added noise. Backpropagation computed gradients for all trainable modules in the network (i.e. SVDFeatureExtractor which processes

frozen SVD components, GAT attention layers, residual blocks, and time embedding layers). In order to provide stability to training and eliminate the occurrence of gradient explosion, gradient clipping was implemented using a maximum norm of 1.0, prior to every optimizer step. Dropout regularization (probability: 5%) was constantly used in training in residual blocks and graph attention layers to improve the generalization of models and to avoid overfitting. Also, GPU cache was periodically cleared every 10 epochs to ensure the efficiency of computation during long training periods. All the training processes were implemented on GPUs using PyTorch with the epoch-wise loss tracking and learning rate change visualization to provide appropriate convergence behavior.

4.1.5 Signal Reconstruction and Evaluation

The reconstruction power of the trained model was tested on a masking-based inpainting process, which was only used during the inference stage, and therefore the model had never seen masked signals during training. As part of evaluation, test EEG segments were applied to a controlled masking strategy: in a temporal (timesteps) window of 180 to 400 (1.4 to 3.1 seconds) each channel approximately 35% was randomly masked with a probability of 70 percent to introduce different patterns of missing data without interrupting contextual information of other unmasked regions. During inference the precomputed SVD components (U, S, V^T) associated with each test segment is passed to the model along with the masked signal, and thus, it was possible to provide the SVDFeatureExtractor with explicit structural (spatial and temporal) priors to guide the reconstruction process to physiologically plausible solutions. The reconstruction process was based on learned reverse diffusion dynamics, which started with masked regions initialized to pure Gaussian noise while leaving known values in unmasked regions. The model generated progressively denoised estimates, running over 300 steps of iterative denoising, moving backward in time from timestep 299 to 0, with U-Net-GAT-SVD architecture utilizing both the learned diffusion dynamics and the mathematical priors by the SVD decomposition to successively estimate and eliminate noise at each step, reintroducing the known unmasked data after each denoising step to maintain signal consistency and provide the denoising process with the benefits of space-time consistency provided by the skip connections of the U-Net and the inter-channel attention processes of the GAT operating on SVD enhanced fused features.
The quality of reconstruction was evaluated thoroughly, based on five complementary measures that were calculated over the masked areas only:

- The point-wise reconstruction accuracy was measured by Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).
- Mean Absolute Error (MAE) was a measure of the magnitude of deviation on average.
- Pearson Correlation Coefficient (PCC) was used to measure the morphology preservation of temporal waveform and phase relationships.

Power Spectral Density (PSD) Distance–the fidelity of frequency-domain properties was measured using Welch method with 64-point (block-long) segments so that the

characteristics of reconstructed signals preserved both the time and spectral dynamics that are neurophysiologically interpretable.

To have strong statistical approximations, all metrics were consolidated across various channels and test samples to obtain the reported values of the mean and standard deviation, which could be used to assess the ability of the model to produce physiologically plausible EEG signals when partial observations were available.

4.2 Model Design Specification

4.2.1 Comparison Among Existing Solutions

The samples produced by generative adversarial networks (GANs) are often of superior visual clarity, but their training is often unstable. Instead, variational autoencoders (VAEs) are prone to over-smooth outputs and thus overlook important high-frequency data.Transformer-based models although able to model long-range interdependencies are computationally demanding and require large data sets to be available.
The current study proves that the denoising diffusion probabilistic model (DDPM) structure is an even better methodological decision. The model that is proposed in this paper will provide high-fidelity and diverse samples as a result of learning to undo a progressive noising process. In such a way, it prevents the traps of other paradigms of generative approaches. The effectiveness of this method is supported by the fact which we have seen from the literature review that DDPM models tend to have better accuracy. These findings suggest that the DDPM approach not only precisely reconstructs corrupt signal sections but retains their critical time dynamics, which makes the approach particularly appropriate in improving EEG analysis.

4.2.2 Denoising Diffusion Probabilistic Model (DDPM)

A customized Denoising Diffusion Probabilistic Model (DDPM) that aims at generating and reconstructing EEG was used. The 1D U-Net backbone with GAT is complemented with self-attention modules and residual connections to take into account local features and global contextual information which multi-channel EEG signals possess. In particular, the architecture is made to mimic a pre-specified diffusion pathway and consequently, the underlying EEG distribution can be modeled and new data can be reconstructed or missing data can be filled in with high fidelity.

Diffusion Process Details

1. **Forward Process (Noising)**:
   The EEG signal x_0 is gradually corrupted with Gaussian noise over T timesteps:

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}\, x_{t-1},\, \beta_t \mathbf{I}\right), \tag{4.2}$$

   where β_t follows a linear schedule from 10^{-4} to 0.02.

Table 4.1: Comparison of Generative Models

Model	**Key Strength**	**Key Weakness**	**Chosen?**	**Reason**
GANs	Generates sharp, high-fidelity samples.	Unstable training and potential for mode collapse.	No	Unreliable training process.
VAEs	Stable training and meaningful data representation.	Blurry outputs.	No	Fails to capture EEG fidelity.
Transformers	Great for long-range dependencies.	Requires high resources.	No	Resource-heavy.
DDPMs	High-quality and stable output.	Slow sampling.	Yes	Best for coherent signal reconstruction.

2. **Reverse Process (Denoising)**:
   The U-Net ϵ_θ predicts noise ϵ at each timestep t, iteratively refining x_t toward clean data:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right), \Sigma_\theta\right), \tag{4.3}$$

   where:

   - $\alpha_t = 1 - \beta_t$
   - $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$
   - x_t: Noisy signal at timestep t
   - x_{t-1}: Denoised signal at timestep $t-1$
   - $\epsilon_\theta(x_t, t)$: Predicted noise by the model
   - β_t: Variance schedule controlling the noise level
   - Σ_θ: Learned or fixed variance for sampling
   - $\mathcal{N}(\mu, \Sigma)$: Gaussian distribution with mean μ and covariance Σ

Model Architecture: 1D U-Net with SVD Decomposition and Graph Attention Networks

The proposed model uses a symmetric encoder-decoder U-Net architecture enhanced with Singular Value Decomposition (SVD) preprocessing and Graph Attention Networks (GAT) to model both temporal dynamics and spatial relationships between

electrode locations. EEG signals of size 32 channels $\times 640$ timesteps are fed through a model block with 32 input/output terminals and 96 model channels. Each stage has three residual blocks with channel multipliers $[1, 2, 4]$, resulting in the downsampling path having feature maps of 96, 192, and 384 channels. Signal integrity at the reconstruction level is maintained by a conservative dropout rate of 0.05.

SVD Precomputation (Offline Preprocessing)

Far from the other way, where the SVD computation is carried out during the learning process, this implementation requires the computation of SVD as a one-time offline learning preprocessing step. For each EEG segment $X \in R^{32\times640}$ in the dataset, SVD is computed once and cached:
Preprocessing Stage (this is done once when the data is loaded).

This method of pre-computation offers:

- 3-4x training speedup (zero redundant matrix factors)
- Features of consistency over time during training
- fewer memory operations during forward/backward pass

During training and inference, the model receives both the original signal X and its precomputed SVD components (U_K, Σ_K, V_K^T) from the cache.

SVD-Based Feature Extraction

Each EEG segment $X \in R^{32\times640}$ is decomposed using Singular Value Decomposition:

$$X = U\Sigma V^T \tag{4.4}$$

where:

- $U \in R^{32\times32}$ contains the left singular vectors (spatial patterns across channels)
- $\Sigma \in R^{32}$ is a vector of singular values (importance weights) (stored as a vector rather than diagonal matrix in implementation)
- $V^T \in R^{32\times640}$ contains the right singular vectors (temporal patterns)

In order to simplify the computation challenge and only to obtain dominant patterns, the decomposition is truncated to the top $K = 16$ components:

$$X \approx U_K \Sigma_K V_K^T \tag{4.5}$$

where $U_K \in R^{32\times16}$, $\Sigma_K \in R^{16}$, and $V_K^T \in R^{16\times640}$.

SVD Feature Processing

The three SVD components are processed separately and then fused:
1. Spatial Pattern Processing (from $\mathbf{U}_K$):

$$\mathbf{h}_{\text{spatial}} = \text{SiLU}(\text{GroupNorm}(\text{Conv1D}(\mathbf{U}_K^T))) \tag{4.6}$$

2. Temporal Pattern Processing (from $\mathbf{V}_K^T$):

$$\mathbf{h}_{\text{temporal}} = \text{SiLU}(\text{GroupNorm}(\text{Conv1D}(\mathbf{V}_K^T))) \tag{4.7}$$

3. Importance Weighting (from $\mathbf{\Sigma}_K$):

$$\mathbf{w}_{\text{importance}} = \sigma(\text{MLP}(\mathbf{\Sigma}_K)) \tag{4.8}$$

where $\sigma(\cdot)$ is the sigmoid activation ensuring weights are in $[0, 1]$.

4. Feature Fusion:

$$\mathbf{h}_{\text{SVD}} = \mathbf{w}_{\text{importance}} \odot [\mathbf{h}_{\text{spatial}} \oplus \mathbf{h}_{\text{temporal}}] \tag{4.9}$$

where $\oplus$ denotes channel-wise concatenation and $\odot$ is element-wise multiplication. This SVD-derived feature representation $\mathbf{h}_{\text{SVD}} \in R^{96\times640}$ is then fused with the original signal features at the input stage of the U-Net.

Graph Attention for Spatial Processing

The 32 EEG channels form a fully connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $|\mathcal{V}| = 32$. In the implementation, GAT is applied in the temporal dimension (with 640 timesteps being nodes with 32 channels features), which indirectly models the inter-channel spatial relationships via attention-weighted features aggregation.
For each channel pair (i, j), attention coefficients are computed as:

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]\right) \tag{4.10}$$

where $\mathbf{W} \in R^{d_{\text{out}}\times d_{\text{in}}}$ is a learnable weight matrix and $\mathbf{a} \in R^{2d_{\text{out}}}$ is the attention parameter.
Normalized attention weights are obtained via softmax:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{32} \exp(e_{ik})} \tag{4.11}$$

Features are aggregated using multi-head attention ($K = 4$ heads):

$$\mathbf{h}_i' = \text{LayerNorm}\left(\Big\|_{k=1}^{4} \sum_{j=1}^{32} \alpha_{ij}^k \mathbf{W}^k \mathbf{h}_j\right) \tag{4.12}$$

Enhanced Input Processing with SVD Integration

The model receives both the original EEG signal $\mathbf{x}$ and its precomputed SVD components $(\mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T)$. The initial feature representation is created by:
1. Original Signal Projection:

$$\mathbf{h}_{\text{orig}} = \text{Conv1D}(\mathbf{x}) \tag{4.13}$$

2. SVD Feature Extraction:

$$\mathbf{h}_{\text{SVD}} = \text{SVDExtractor}(\mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T) \tag{4.14}$$

3. Feature Fusion:

$$\mathbf{h}_0 = \text{SiLU}\Big(\text{GroupNorm}\big(\text{Conv1D}([\mathbf{h}_{\text{orig}} \oplus \mathbf{h}_{\text{SVD}}])\big)\Big) \tag{4.15}$$

This fused representation $\mathbf{h}_0$ serves as the input to the encoder, incorporating both raw temporal information and low-rank structural patterns extracted by SVD.

Encoder Blocks with GAT Integration

Each encoder block fuses temporal and spatial features. Temporal processing uses residual blocks:

$$\mathbf{h}_{\text{temp}} = \text{ResBlock}_3\big(\text{ResBlock}_2\big(\text{ResBlock}_1(\mathbf{x}, \mathbf{t}_{\text{emb}})\big)\big) \tag{4.16}$$

where $\mathbf{t}_{\text{emb}}$ is the timestep embedding in the diffusion process.
Spatial processing applies GAT with learned projection:

$$\mathbf{h}_{\text{spatial}} = \text{Proj}\big(\text{GAT}(\mathbf{h}_{\text{temp}})\big) \tag{4.17}$$

Adaptive fusion combines both features:

$$\mathbf{h}_{\text{out}} = \sigma(\lambda) \cdot \mathbf{h}_{\text{temp}} + (1 - \sigma(\lambda)) \cdot \mathbf{h}_{\text{spatial}} \tag{4.18}$$

where λ is a learnable parameter and $\sigma(\cdot)$ is the sigmoid function.

Network Components

Downsampling: Strided 1D convolutions (kernel 3, stride 2) reduce temporal resolution.
Upsampling: Transposed 1D convolutions (kernel 4, stride 2) with skip connections restore resolution.
Residual Blocks: Use group normalization (8 groups), SiLU activation, and time embedding injection between convolution layers. The residual block operation follows:

$$\begin{aligned}
\mathbf{h}_1 &= \text{SiLU}(\text{GroupNorm}(\text{Conv1D}_1(\mathbf{h}))) \\
\mathbf{h}_2 &= \mathbf{h}_1 + \text{MLP}(\mathbf{t}_{\text{emb}}) \\
\mathbf{h}_3 &= \text{SiLU}(\text{Dropout}(\text{GroupNorm}(\text{Conv1D}_2(\mathbf{h}_2)))) \\
\mathbf{h}_{\text{out}} &= \mathbf{h}_3 + \text{Shortcut}(\mathbf{h})
\end{aligned} \tag{4.19}$$

where the time embedding is injected after the first convolution layer and before the second convolution layer.
Time Embedding: Sinusoidal encoding of 96 dimensions:

$$\mathbf{t}_{\text{emb}}[2i] = \sin\left(\frac{t}{10000^{2i/96}}\right) \tag{4.20}$$

$$\mathbf{t}_{\text{emb}}[2i+1] = \cos\left(\frac{t}{10000^{2i/96}}\right) \tag{4.21}$$

This base 96-dimensional sinusoidal embedding is then projected to 384 dimensions via a two-layer MLP:

$$\mathbf{t}_{\text{emb}}^{384} = \text{MLP}(\mathbf{t}_{\text{emb}}^{96}) \tag{4.22}$$

Diffusion Framework

The model operates within a Denoising Diffusion Probabilistic Model (DDPM) with 300 timesteps and a linear noise schedule $\beta_t \in [0.0001, 0.02]$.
Forward Diffusion Process:
At each timestep t, Gaussian noise is progressively added to the clean signal $\mathbf{x}_0$:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \tag{4.23}$$

where

$$\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s) \tag{4.24}$$

is the cumulative product of noise scheduling coefficients.

Usage of Precomputed SVD in Diffusion

The precomputed SVD components serve as auxiliary conditioning information throughout the diffusion process:

Training Phase:
For each training iteration, the model receives:

1. Noisy signal: $\mathbf{x}_t$ (corrupted with noise at timestep t)
2. Diffusion timestep: $t \in \{0, 1, ..., 299\}$
3. Precomputed SVD components (from cache): $\mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T$

The network predicts the noise conditioned on all three inputs:

$$\hat{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T) \tag{4.25}$$

Loss Function:

$$\mathcal{L}_{\text{simple}} = E_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T)\|^2\right] \tag{4.26}$$

Critical Note: The SVD components $(\mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T)$ are computed from the clean signal $\mathbf{x}_0$ during preprocessing and remain constant throughout training—they are never recomputed for noisy versions $\mathbf{x}_t$. This provides stable structural priors that guide the denoising process.

Reverse Diffusion Process (Sampling/Reconstruction):

During inference, the reverse process iteratively denoises from pure noise $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$ back to clean signal $\mathbf{x}_0$:
At each reverse step $t \to t-1$:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1-\beta_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{U}_K, \mathbf{\Sigma}_K, \mathbf{V}_K^T) \right) + \sigma_t \mathbf{z} \tag{4.27}$$

where:

- $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ is sampled noise
- $\sigma_t = \sqrt{\beta_t \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}}$ is the posterior variance

Key Implementation Detail: At every reverse diffusion step (from $t = 299$ down to $t = 0$), the same precomputed SVD components (derived from the original clean signal or known regions) are passed to the network. This provides consistent structural guidance throughout the entire denoising trajectory.

Masked Reconstruction with SVD Conditioning:

For EEG signal reconstruction with masked regions, the diffusion process preserves known data at each reverse step:

$$\mathbf{x}_{t-1} = \mathbf{x}_{t-1}^{\text{predicted}} \odot (1 - \mathbf{M}) + \mathbf{x}_{\text{known}} \odot \mathbf{M} \tag{4.28}$$

where:

- $\mathbf{M} \in \{0, 1\}^{32 \times 640}$ is the binary mask (1 = known, 0 = unknown)
- $\mathbf{x}_{\text{known}}$ contains the observed EEG values
- $\mathbf{x}_{t-1}^{\text{predicted}}$ is the denoised prediction from the model

This iterative process that leaves the masked EEG regions open for reconstruction, conditioned on precomputed SVD features at each iteration, provides high fidelity reconstruction of masked EEG regions without distorting the underlying structure in space and time captured by the low-rank decomposition process.

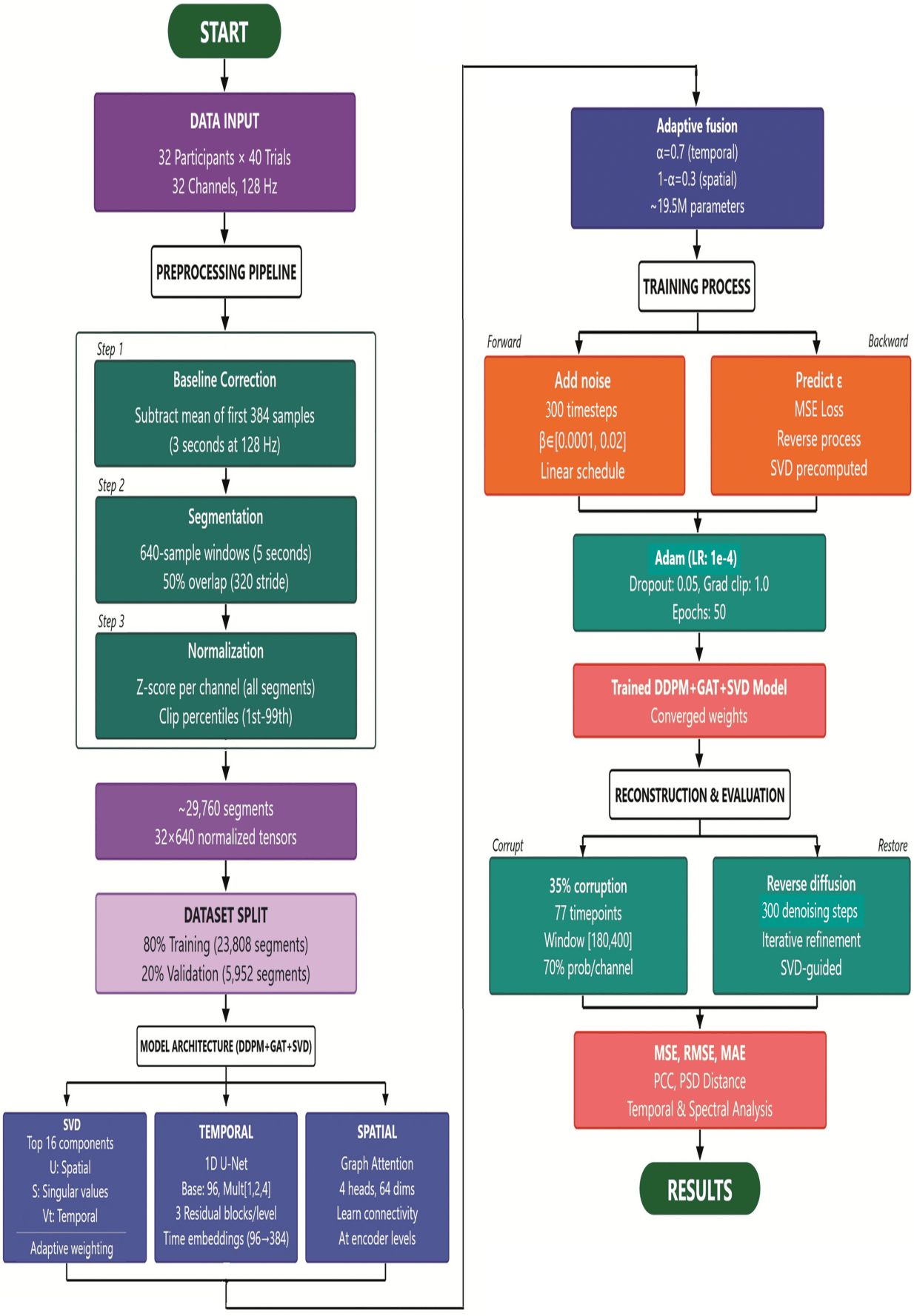


Figure 4.3: Proposed Model Architecture

4.3 Data Collection

The paper uses the DEAP [2] (Dataset for Emotion Analysis using Physiological Signals) dataset that consists of EEG data of 32 subjects who completed 40 trials of 32 channels at a sampling frequency of 128Hz. The DEAP dataset used for training and primary evaluation was obtained from Kaggle though we have applied for the original DEAP dataset from their main website but unfortunately got no reply from them. The DEAP dataset was chosen because it has a balanced set of standardized music video stimuli, a rich 32-channel EEG signal, and an adequate sample size (32 participants x 40 trials). That is why it is especially suitable to be used in our study of emotion-based generation of EEG signals. Other effective computing data sets that are popular were however ruled out after being looked at. SEED dataset, which was supplied by Shanghai Jiao Tong University, provides high-density EEG data (62 channels) and is quite similar to our intention to focus on the dimensional model of emotion, although still limited to a small number of participants (15) and discrete category of emotion (positive/neutral/negative). In the same vein, the MAHNOB-HCI dataset, with its multimodal EEG, physiological signals, and facial video records of 30 subjects, is also problematic because of its reduced number of trials (20 per subject) and the lack of consistency in the emotion induction procedure with mixed stimuli (images and videos). Although the DREAMER dataset included EEG and ECG data of 23 subjects, it was deemed to be less appropriate because of the low spatial resolution (14-channel Emotiv EPOC system), and possible signal quality issues related to consumer-grade equipment. Lastly, the AMIGOS dataset has a wider population (40 participants) and contextual social background, however, its emphasis on group interactions and 14-channel EEG afforded us fewer options and synergy with our research needs than the higher-density recording and controlled individual emotion induction paradigm in DEAP. PhysioNet Motor Movement/Imagery (109 subjects, 64 channels, variable-length recordings) was not used as the primary training dataset because of the motor imagery task paradigm (inconsistent with valence-focused research objectives), shorter trial lengths represented by longer sequences insufficient to support adequate temporal modeling, and substantial additional preprocessing required that would have included many potential confounders in our evaluation. The dataset was however later used as independent test benchmark in order to verify the cross-dataset and cross-paradigm generalization capabilities, illustrating that the proposed DDPM+GAT+SVD architecture can consistently provide reconstruction performance not sensitive to task domain, channel layout or acquisition protocols and therefore establish universality beyond the DEAP training distribution.

4.3.1 Data Cleaning

Data cleaning step: The baseline correction is done by subtracting the average of the first 3 seconds (384 samples) of each trial in order to remove DC offsets. The outlier mitigation is attained via robust clipping which limits the values to 1st and 99th percentiles, as part of the normalization. One of the weaknesses is that there are no clear artifact removal algorithms as well (e.g., ocular or muscular noise) which could negatively affect the performance of the model on real-world recordings with high levels of artifacts.

4.3.2 Data Transformation

Data transformation is performed by z-score normalization of each of 32 EEG channels independently, making the amplitudes have zero mean and unit variance and minimizing the electrode-dependence. Percentile based clipping is a method for clipping normalized values into the 1st to 99th percentile, so that the values for the extreme outliers are attenuated, but not the extreme segments. This strategy of normalization and clipping helps in the stability of the input distribution and is helpful for the diffusion process modeling by matching data displacement to noise schedule.
EEG signals are divided into windows of 640 samples (5 seconds at 128 Hz in this example) using the sliding window method with 50% overlapping (stride of 320 samples). This length of the window allows for enough temporal context to capture dynamic EEG patterns and at the same time being computationally tractable.
Following normalization and segmentation, Singular Value Decomposition (SVD) is precomputed for each 640-sample window represented as a 32×640 matrix with spatial ($\mathbf{U}$), singular value ($\mathbf{\Sigma}$), and temporal ($\mathbf{V}^T$) components of the signal.
In order to reduce computational cost and suppress the low-energy components, truncated SVD is performed by taking the leading 16 singular components, which yields:

$$\mathbf{U}_{\text{trunc}} \in R^{32\times16}, \quad \mathbf{\Sigma}_{\text{trunc}} \in R^{16}, \quad \mathbf{V}^T_{\text{trunc}} \in R^{16\times640} \tag{4.29}$$

These precomputed components of the SVD are also saved together with the normalized EEG segments and used again during the training process so that the SVD decomposition is not repeated while yielding clear representations of the structure of spatial-temporal patterns.
No temporal smoothing or frequency domain filtering is used, which maintains the physiological integrity of the EEG signals.

4.3.3 Data Integration

Data integration cannot be applied in the study, since the DEAP dataset is self-sufficient, and it does not need to be combined with any external sources. The protocol of the data preprocessing is internally consistent, and no integration of the datasets is necessary. This strategy guarantees homogeneity in the input data, which is essential in training diffusion models not to add confounding variability due to heterogeneous sources.

4.3.4 Data Reduction

The study does not reduce the number of EEG channels, but keeps all the 32 channels to keep spatial resolution because explicit dimensionality reduction methods such as the Principal Component Analysis (PCA) may lose neurophysiologically meaningful spatial information. Temporal downsampling is not attempted to preserve the original sampling rate of 128 Hz and preserve high frequency signal content.
Instead, dimensionality reduction is presented using SVD based rank truncation. Rather than reducing the number of channels or the timesteps, in SVD, each EEG segment is decomposed into a set of orthogonal components, which are spatial/temporal, and the leading 16 components are taken from the set.

This rank reduction is done on the mathematical structure of the signal matrix without reducing the original spatial (32 channels) and temporal (640 timesteps) resolution of the raw data. The resulting compact representation gives focus on spatial-temporal patterns which is dominant while suppresses the low-energy components.
Feature compression is taken care of also implicitly, in the encoder pathway of the U-Net architecture by means of the progressive downsampling of latent feature representations, without affecting the dimensionality of the raw input data in the data preprocessing step.
The selection of 640 samples window length is a good balance between temporal context and computation power, which can optimize the trade-off to obtain stable and effective model training. The combination of full resolution EEG signals and SVD-based rank reduction offers structured priors to guide diffusion models without losing neurophysiological information.

4.3.5 Summary of Preprocessed Data

The data is pre-processed to obtain data in segments with 32 EEG channels. The duration of each segment is 5 seconds (640 samples at 128 Hz). Global per-channel z-score normalisation is applied across the entire dataset, for which the mean (μ) and standard deviation (σ) are calculated for each of the samples of each channel independently. This guarantees zero mean and unit variance per channel, so that all channels could make a similar contribution to help the learning process and to improve the stability of gradient-based optimization.
Following normalization, percentile clipping (1st and 99th percentiles) is performed to reduce the effect of extreme outliers without affecting the integrity of the physiological signals. This generates stable and well-conditioned inputs for consistent and robust learning.
To maximize the utilization of data and help with temporal pattern learning, segmentation is carried out with 50% of overlap (stride of 320 samples). This has two advantages: it increases the number of training samples each trial and enables the model to learn temporal dynamics and variation of emotional response.
Singular Value Decomposition (SVD) is used on each segment and decomposes it into $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}^T$ components. Only the top 16 singular components are preserved, and it offers a low-rank compact representation of the data that can capture the main spatial-temporal patterns and still keeps the computation low-dimensionality. These SVD features are pre-computed and cached offline during the construction of the dataset thus it avoids redundant computations during training.
Descriptive statistics such as mean (μ), standard deviation (σ), minimum and maximum are computed in a channel-by-channel fashion, in order to assure statistical equality throughout the dataset. These metrics are used to validate the effectiveness of the preprocessing pipeline and ensure that the data fits neurophysiological patterns.
Finally, the fully-preprocessed version of the dataset (with the original segments and its precomputed SVD components) is wrapped into a PyTorch Dataset object so that the batch loading with the precomputed features can be performed efficiently during training. Such design is customized according to the model input requirements of the diffusion model while maintaining the physiological fidelity required to properly

recreate and generate emotion-based EEG signals.

4.4 Implementation of Selected Design

The implementation integrates Denoising Diffusion Probabilistic Models (DDPM) with Graph Attention Networks (GAT) and Singular Value Decomposition (SVD) preprocessing to reconstruct corrupted EEG signals through probabilistic refinement. This architecture combines a diffusion model for gradual restoration with graph-based spatial modeling to capture inter-channel interdependencies across the 32-electrode montage while using low-rank decomposition to extract dominant spatial-temporal patterns and minimize computational overhead.

The main architecture consists of a hierarchical backbone containing a 1D U-Net extended through multi-head GAT layers at every encoder stage. The U-Net utilizes 96 base channels expanded through three levels using channel multipliers [1, 2, 4], which form 96-dimensional, 192-dimensional, and 384-dimensional feature representations. At the input part, the network takes both the raw EEG signal and its precomputed SVD components (spatial patterns $\mathbf{U}$, singular values $\mathbf{\Sigma}$, and temporal patterns $\mathbf{V}^T$) as inputs, which are proposed through different convolutional pathways and also fused with the signal raw EEG feature. The spatial patterns from $\mathbf{U}$ are processed using 1D convolutions to extract the channel relationships, temporal patterns from $\mathbf{V}^T$ being the temporal behavior of the signal are indexed from $\mathbf{V}'$ and finally, singular values $\mathbf{\Sigma}$ are transformed using multilayer perceptrons to produce importance weights modulating the contribution of SVD features. This multi-stream fusion is the way by which the model can have both raw temporal information and structural priors that are obtained through low-rank approximation.

Each encoder block contains three residual sub-blocks with GroupNorm and SiLU activation to achieve stable gradient flow. Strided convolutions perform downsampling between levels, progressively compressing the temporal dimension while expanding the feature space to represent increasingly abstract patterns. The integration of GAT treats EEG channels as nodes in a graph, with their interrelationships representing functional connectivity. The GAT layers have 4 attention heads altogether of a total output dimensionality of 64 (16 dimensions per head), with that being projected to match the U-Net channel width at each level of the encoder, dynamically learning attention weights for channel pairs. This is in contrast to fixed anatomical connectivity, which discovers relationships in the data reflecting not only local dependencies (proximity of electrodes) but also long-range trends (bilateral synchronization). GAT outputs are projected to match the U-Net dimensions and combined with the temporal features through an optimized during training and weighted sum which is learnable, that adaptively balances the two modalities during training. The network is also conditioned on diffusion timesteps (dimension 384) which enable the model to adapt denoising strategies—learning coarse structure when noise is high and refining fine details when noise is low. The decoder mirrors the encoder, using transposed convolutions to upsample the signal and skip connections that provide direct pathways across the bottleneck for preserving high-resolution information.

The final encoder projection converts 96-channel features back to 32 EEG channels, predicting the noise component in both forward and reverse diffusion directions. Training follows the standard DDPM framework where clean signals (32 channels $\times$ 640 timepoints at 128 Hz) are corrupted across 300 timesteps using a linear vari-

ance schedule ($\beta \in [0.0001, 0.02]$). For each training iteration, on each timestep (a randomly sampled one) from a batch is the noisy EEG signal, precomputed SVD components ($\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}^T$), retrieved from cache. The SVD components are computed once during the preprocessing from the clean signals and remain constant throughout the training, and act as auxiliary conditioning information. This design avoids redundant matrix factorizations that are computed on every forward pass; consequently training time is significantly shorter than on-the-fly SVD computation. At each iteration, a timestep is randomly sampled and Gaussian noise is added using closed-form expressions of cumulative noise coefficients. The model minimizes the expected error between predicted and actual noise. Adam optimizer with weight decay (1×10^{-4}) is used with learning rate of 1×10^{-4} and gradient clipping (max norm 1.0) and with cosine annealing schedule 50 epochs with batch size 16. The addition of SVD preprocessing modules contributes minimal overhead while providing substantial computational benefits during training.

As a key optimization step in the preprocessing pipeline, SVD decomposition is used. During the loading of the data, the 640 samples of EEG data are subjected to matrix factorization to extract the top 16 singular components. This truncated decomposition captures the dominant modes of variation in both the spatial dimensions (the channels) and the temporal dimensions, and throws away the noise-dominated components of variation. The precomputed triplets ($\mathbf{U}_{16}, \mathbf{\Sigma}_{16}, \mathbf{V}_{16}^T$) are stored along with the original segments in the memory which allows retrieving them instantly during the training without recomputation. This one-time preprocessing takes a few minutes on the entire DEAP dataset, and provides huge efficiency improvements during the training process.

For reconstruction, the masked inpainting approach introduces an approximation of realistic artifacts by removing 35% of the signal (approximately 77 consecutive timepoints) within the window [180, 400], which has been selected to avoid edge effects while ensuring sufficient surrounding context. Reconstruction begins with the masked input where corrupted regions are replaced with Gaussian noise. The model then performs reverse diffusion for 300 steps such that during each timestep, the current noisy signal, the index of the timestep and the precomputed SVD components from the original clean signal are fed into the timestep. These SVD features provide structural guidance in the entire trajectory of the denoising procedure, and they help the model to produce physiologically correct reconstructions and respect the underlying patterns of the space-time based on the low-rank decomposition. At each step, the signal is refined by predicting and subtracting noise according to posterior variance equations. Critically, at every step, known signal segments are clamped to their original values, ensuring that only corrupted regions are refined. GAT layers leverage learned attention to propagate information from intact channels to masked regions, exploiting inter-channel relationships to construct physiologically plausible patterns based on spatial dependencies. The combination of diffusion-based iterative refinement, graph attention for spatial coherence, and structural priors derived from SVD allows achieving robust reconstruction while preserving temporal continuity as well as cross-channel consistency.

Preprocessing of inputs includes baseline correction by subtracting average levels over the first 384 samples and channel-wise z-score normalization (mean $\mu = 0$, standard deviation $\sigma = 1$) per channel, which accounts for amplitude variation across locations, followed by percentile clipping to mitigate outliers without removing en-

tire segments. Signals are segmented into 640-sample windows with 50% overlap (320-sample stride), which increases training samples and avoids boundary information loss while enhancing phase-invariant learning. Following the segmentation and the normalization, it is applied to each of the segments that were extracted: SVD decomposition, extracting the first 16 components to form compact feature representations. These precomputed SVD features are packaged along with their respective segments into PyTorch tensors, and they are efficiently loaded during training using custom collate functions. This results in about 29,760 segments from the DEAP dataset after preprocessing. The dataset is free of synthetic augmentation (noise injection, phase shifts and amplitude perturbations) and only uses natural variability and 35% masking ratio during training data to ensure faithful replication of neurophysiology original properties.

All hyperparameters (300 timesteps, noise schedule $\beta \in [0.0001, 0.02]$, masking ratio of 35%, GAT architecture (4 heads, 64 dimensions) and SVD of 16 components) are determined through some preliminary experimental studies. Using the SVD precomputed pipeline we have finished the training in 50 epochs efficiently on one GPU with checkpoints stored every 10 epochs during training. This offers a closed and reproducible system for proving the practicality of implementing advanced diffusion models augmented with low-rank decomposition for neurophysiological signal reconstruction. The integration of SVD not only speeds up the training of the model but also adds knowledge of explicit structural information to the model, which makes the model learn more efficiently from the natural spatial-temporal organization of EEG data.

Chapter 5

Result Analysis

5.1 Performance Evaluation

5.1.1 Evaluation Metrics

Five measures for quantitative evaluation were used to evaluate the quality of EEG signal reconstruction in masked areas comprehensively. These measures assess various reconstruction fidelity measures such as accuracy pointwise, preservation of temporal patterns and spectral performance. All measures achieved were calculated only considering the corrupted regions (the corrupted regions were about 77 consecutive timepoints randomly selected in the window [180, 400] which is 35 percent of the availed masking window) to test the reconstruction capacity of the model and not the recall ability to preserve known information of the model.

Mean Squared Error (MSE) Mean Squared error measures the mean squared error between the original masked signal x and the reconstructed signal o x in the masked region:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{5.1}$$

N is used in place of masked time points. MSE penalizes larger errors In quadratic terms and thus it is especially sensitive to outliers and important reconstruc- tion failures. When it comes to the EEG reconstruction, the MSE values less than 0.03 suggest. high quality of reconstruction, and the values above 0.08 indicate poor performance in the values of performance. This measure is used as the major optimization goal during model training.

Root Mean Squared Error (RMSE) RMSE provides error magnitude in the same units as the original EEG signal (microvolts after normalization), computed as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{5.2}$$

In comparison with MSE, RMSE is easier to interpret because it simply states the characteristic amplitude deviation between original and reconstructed signal(s).

Values below 0.20 are discussed to be permissible to normalized EEG data, with values below 0.15 indicating high-fidelity reconstruction.

Mean Absolute Error (MAE) MAE is the mean(average) of the original and reconstructed absolute deviation signals:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i| \tag{5.3}$$

As compared to MSE, MAE uses linear penalty to all the errors, no matter how small they are, then make it less susceptible to the outliers. This metric gives a moderate evaluation of reconstruction accuracy that does not place too much emphasis on big deviations. For acceptable reconstruction of normalised EEG signals,MAE values is less than 0.15, whereas values less than 0.10 indicates excellent performance.

Pearson correlation coefficient (PCC). PCC evaluates the preservation of temporal patterns and phase relationships between original and reconstructed signals:

$$\text{PCC} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(\hat{x}_i - \bar{\hat{x}})^2}} \tag{5.4}$$

where $\bar{x}$ and $\bar{\hat{x}}$ are the mean values of data of the original and reconstructed signals respectively. PCC is computed that falls between -1 to +1 where smaller values like closer to +1 represent the stronger linear correlation. This measure is of especial importance to EEG reconstruction, because it determines whether the temporal dynamics in the reconstructed signal have been maintained or not and oscillatory patterns characteristic of neural activity. When PCC exceeds 0.80 then there is strong correlation but beyond 0.85 is excellent temporal fidelity. PCC is robust against linear scaling unlike point-wise error metrics to amplitude variations and emphasizing on waveform shape preservation.

Power Spectral Density Distance (PSD Distance) PSD Distance quantifies the discrepancy between the frequency content of original and reconstructed signals:

$$\text{PSD Distance} = \sqrt{\frac{1}{M}\sum_{f=1}^{M}(P_x(f) - P_{\hat{x}}(f))^2} \tag{5.5}$$

Here Px(f) and P^x(f) indicates the power spectral densities of the original and frequency f which is reconstructed signals and M is the frequency bins. PSD was calculated on the basis of Welch method at the correct length of the segment (nperseg = min(64, N/2)) in order to balance frequency resolution vs. variance. This metric is essential EEG: to analyze the EEG data, the neural oscillations within certain frequency bands are required (delta: 0.5-4).Hz, theta: 4-8 Hz, alpha: 8-13 Hz, beta: 13-30 Hz, gamma: 30 Hz) the loadings have different values of physiological and cognitive significance. Values that are lower in PSD Distance are an indication of better preservation of spectral characteristics, where values less than 0.01, are regarded as excellent for normalized signals.

Metric Selection Rationale The choice of the five complementary measures is based on the fact that numerous dimensions of reconstruction quality are being examined. MSE, RMSE, and MAE enable an evaluation of how accurately the various sensitivity profiles are to outliers. PCC examines temporal forms that are not dependent on amplitude scaling, that is essential to keep up the neurophysiological validity of reconstructions of signals. PSD Distance makes sure that the characteristics of frequencies, which are fundamental to EEG interpretations are properly recorded. A combination of these measures will give a continued, multi-dimensional evaluation system that incorporates both temporal and frequency-domain reconstruction fidelity, which allows us to comprehensively validate the architecture of DDPM+GAT+SVD.

Table 5.1: Performance Thresholds for Evaluation Metrics

Metric	**Excellent**	**Good**	**Fair**	**Poor**
MSE (↓)	< 0.02	$0.02 - 0.03$	$0.03 - 0.08$	> 0.08
RMSE (↓)	< 0.15	$0.15 - 0.20$	$0.20 - 0.25$	> 0.25
MAE (↓)	< 0.10	$0.10 - 0.15$	$0.15 - 0.20$	> 0.20
PCC (↑)	> 0.85	$0.80 - 0.85$	$0.70 - 0.80$	< 0.70
PSD Distance (↓)	< 0.01	$0.01 - 0.02$	$0.02 - 0.03$	> 0.03

5.1.2 Testing Methodology

To test the performance of the trained DDPM+GAT+SVD model with respect to the reconstruction of EEG in controlled masking conditions on the DEAP dataset, the trained models were tested. All the experiments were performed on the Kaggle platform with an NVIDIA Tesla P100 with 16 GB of memory and running PyTorch 1.13 with support of CUDA 11.6. Despite being self-supervising in nature, an 80–20 random segment-level traintest split was used in order to ascertain that the performance of reconstruction was measured on unseen signal segments within the same data sample. Assessment was accomplished by both qualitative and visual measurement of five randomly chosen signal segments and quantitative measurement of the entire set of tests, which had approximately 6,000 EEG segments. The stochastic masking with the evaluation was done with randomized patterns that were not in the training stage and in training the model learned the denoising by means of the common diffusion noise process. This architecture is reasonable, since, while learning the underlying structure of EEG signals is the aim of the task, no prediction or classification is required, the stochastic masking is a structured corruption of the signal, which can never be compared with the Gaussian noise used to train and the main objective was to demonstrate the structural capability of the DDPM+GAT+SVD framework, not to provide certain guarantees of robust generalization.
During testing, explicit masking was used to simulate signal corruption. In each of these segments, the fraction of the temporal sub-window $[180, 400]$ (≈ 77 consecutive timepoints out of the 220-point window) chosen that is randomly masked was approximately 35% and the initial location within that temporal sub-window was randomized per sample. Each channel was masked with a probability of 70% channel and had an approximation of 70% percent masked areas per segment with the rest of the channels being not masked. The channel-wise stochastic masking

creates realistic artifact conditions by varying channels and is designed to enable the use of partial spatial information to reconstruct the missing signal by relying on GAT-based inter-channel dependence.
The reconstruction was done with 300-steps reverse diffusion, beginning at $T = 300$ then gradually decreasing to $T = 0$. The U-Net+GAT+SVD learned the noise at each step and then eliminated the noise with the DDPM posterior sampling equation, making use of the precomputed SVD features ($U \in R^{32\times16}, S \in R^{16}, V^T \in R^{16\times640}$) to benefit the capabilities of the model to represent spatiotemporal patterns. The SVD components which have also been precomputed in the data preprocessing phase were inputted into specific convolutional processing modules in every step of the denoising process to enhance the latent representations with decomposed spatial and time information. Those parts of the signal which were uncovered at each timestep were forced to its initial value, so that only corrupted values were adjusted and so as to be spatially and temporally consistent with the observed data. Inter-channel information flow was dynamically modulated in the GAT module at every denoising step through the propagation of patterns across intact channels to masked areas through learned attention weights whereas the features obtained by the SVD offered complementary dimensional reduction and pattern extraction.
All the testing has been done on Kaggle hardware which comprises an NVIDIA Tesla P100 card (16 GB), an Intel Xeon processor, 30GB system RAM, and Ubuntu 20.04. It was based on the software stack of PyTorch 1.13 using CUDA 11.6, NumPy 1.23, SciPy 1.9, Scikit-learn 1.1, and Matplotlib 3.5. The full pipeline of evaluation, such as loading the data, SVD preprocessing, model inference and computing the metric all were contained in a single Python script that was run in a Jupyter notebook setup. The quantitative assessment was performed in a single evaluation without repetition of initializations of the experiment. In every test segment, the 300-step reverse diffusion reconstructed signals were compared to the original ground-truth signals using five measurements (MSE, RMSE, MAE, PCC, and PSD Distance), however, each was only done on the masked regions in order to isolate reconstruction performance. Measurements were made on individual channel and segment basis and the reported values are the average and standard deviation of measurements made on the entire sample of masked. This single-run critique approach was taken in to evaluate reconstruction performance based on a fixed configuration as opposed to being sensitive to randomized set up.
Lack of a participant-level held-out test set is a weakness in methodology. Although the use of segment-level train–test splitting and stochastic masking is an operation that offers partial results in terms of generalization, in future studies, participant-level k-fold cross-validation and testing on external EEG datasets including PhysioNet and BCI Competition data should be used to reinforce the assertion about generalization and cross-domain transferability.

5.2 Analysis of Design Solutions

The suggested approach makes use of an augmented Denoising Diffusion Probabilistic Model (DDPM) that is specifically designed to work with EEG signal reconstruction using a 1D U-Net network. The model was developed and tested on a large scale on the DEAP dataset, comprising nearly 29,760 EEG segments (after 80/20 train-test split). Three architectural versions were implemented and evaluated to

determine the effect that SVD decomposition and graph attention mechanism had on reconstruction performance: (1) DDPM+GAT+SVD (proposed full model), (2) DDPM+GAT (ablation without SVD), and (3) DDPM+SVD (ablation without GAT). The DDPM+GAT+SVD model was chosen as the ultimate implementation because it performs better in all the evaluation criteria.

5.2.1 Model Architectures and Design Differences

- **Model 1: DDPM+GAT+SVD (Proposed Full Model):**
  - The SVDFeatureExtractor module is the key component, with a 4-head multi-head Graph Attention Network (GAT), and an encoder–decoder that has 3 levels ($96 \rightarrow 192 \rightarrow 384$ channels at bottleneck) and channel multipliers equal to $[1, 2, 4]$.
  - SVD integration: Given explicit spatial–temporal decomposition with 16 components truncated (truncated SVD, 16 of 32 components) mathematically represented by U matrices (32×16) of spatial patterns and $V^\top$ matrices (16×640) of temporal patterns via 1D convolution with kernel size 3, approximation by: S vectors (16) are weighted by a 2-layer MLP. The raw signal features (96 channels) are combined with SVD features (96 channels) through 1×1 convolution in order to obtain combined representation ($192 \rightarrow 96$ channels).
  - GAT spatial attention: The attention heads (4) have 64 hidden dimensions that act on post-fusion latent features and any EEG channel (32) is a node in the graph and has learned attention weights that rapidly represent functional connectivity but do not explicitly represent adjacency matrices.
  - Strategy normalization: Z-score normalization channel by channel (mean $\mu = 0$, standard deviation $\sigma = 1$) and 1st–99th percentile clipping, which leaves the range of normalized values around $[-2.58, 2.58]$ standard deviations.
  - To complete the whole training process, a maximum of 50 epochs were needed using a learning rate schedule ($\text{lr} = 1 \times 10^{-4}$), pointing to an approximation of 27.2% reduction in original training loss.
  - Computational efficiency: as the SVD was precomputed, it helps to reduce per-iteration decomposition overhead and the training time per epoch.
  - Achieves a final result with MSE of 0.014153 ± 0.014911, RMSE of 0.110374 ± 0.044388, MAE of 0.087305 ± 0.034363, PCC of 0.8766 ± 0.1207, and PSD error of 0.001293 ± 0.003648.

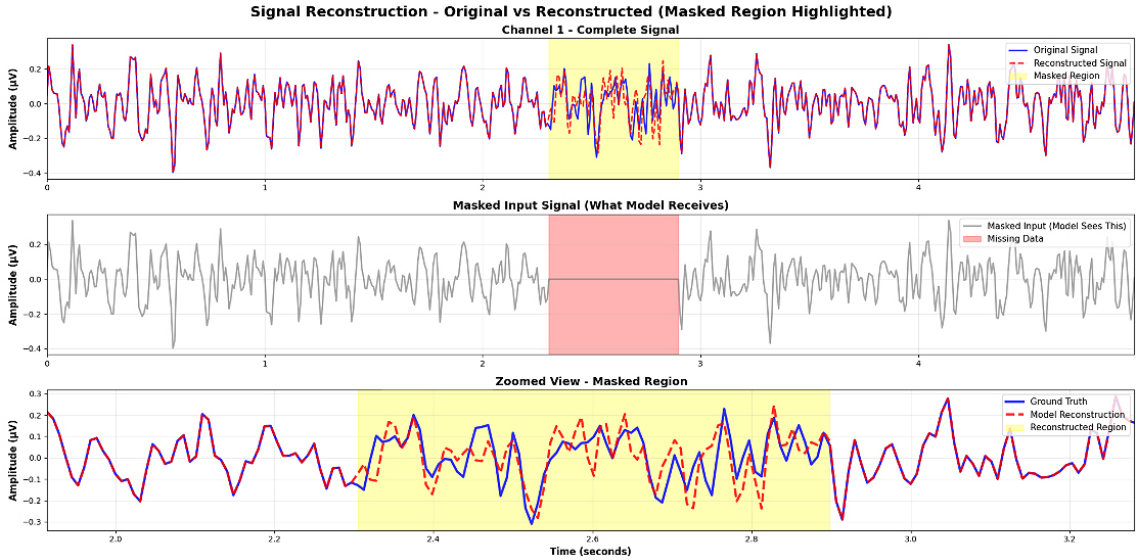


Figure 5.1: DDPM+GAT+SVD Model EEG Signal Reconstruction Output

- **Model 2: DDPM+GAT (Ablation Study – No SVD):**
  - Key components: 4-head GAT with 64 hidden dimensions (this is faster); encoder–decoder levels are different yet four channel multipliers of the same value $[1, 2, 4]$, no SVD integration.
  - Architecture variance: Lacks SVDFeatureExtractor module and fusion layer; raw 32-channel input is processed by initial convolution ($32 \rightarrow 96$ channels) without mathematical decomposition priors.
  - GAT acts only on learned attributes of the raw signals without any guidance by SVD in the form of spatial–temporal structure.
  - Adopts the same normalization approach as Model 1 (channel-wise z-score + percentile clipping).
  - Model training for 50 epochs using the same optimizer settings (Adam, $\text{lr} = 1 \times 10^{-4}$, cosine annealing).
  - Limitation of performance: Without SVD's explicit spatial–temporal decomposition, the model uses learned representations solely to generate reconstruction. The performance of this model is not as high as that of the full model, showing that mathematical priors (SVD) play a critical role in the diffusion process.

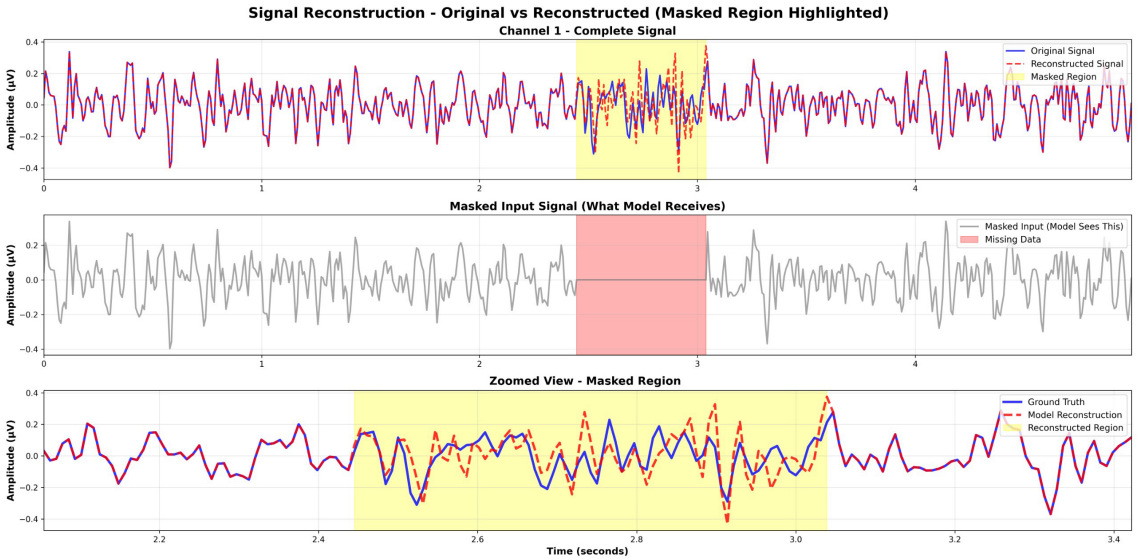


Figure 5.2: DDPM+GAT Model EEG Signal Reconstruction Output

- **Model 3: DDPM+SVD (Ablation Study – No GAT):**
  - Key components: SVD processor with 16-component truncation, standard residual blocks without graph attention, 3-level encoder–decoder with channel multipliers $[1, 2, 4]$.
  - Architecture difference: It has both SVD decomposition and precomputation; it does not have the adaptive spatial attention feature of GAT. Encoder blocks only contain residual convolutions followed by group normalization and SiLU activation, and do not have a representation of inter-channel dependency.
  - GAT is not supported by SVD, which offers spatial–temporal decomposition (U, S, $V^\top$ components), but it lacks the ability to learn adaptive attention weights between channels of EEG activity capturing different brain regions.
  - All other parameters are the same as in Models 1 and 2.
  - Performance limitation: Although SVD offers mathematical structure, the absence of GAT means no spatial attention is learned; thus, no dynamic inter-channel relationships and functional connectivity patterns of EEG reconstruction can be realized by the model. This shows that mathematical priors (SVD) and learned attention (GAT) are both required to maximize performance.

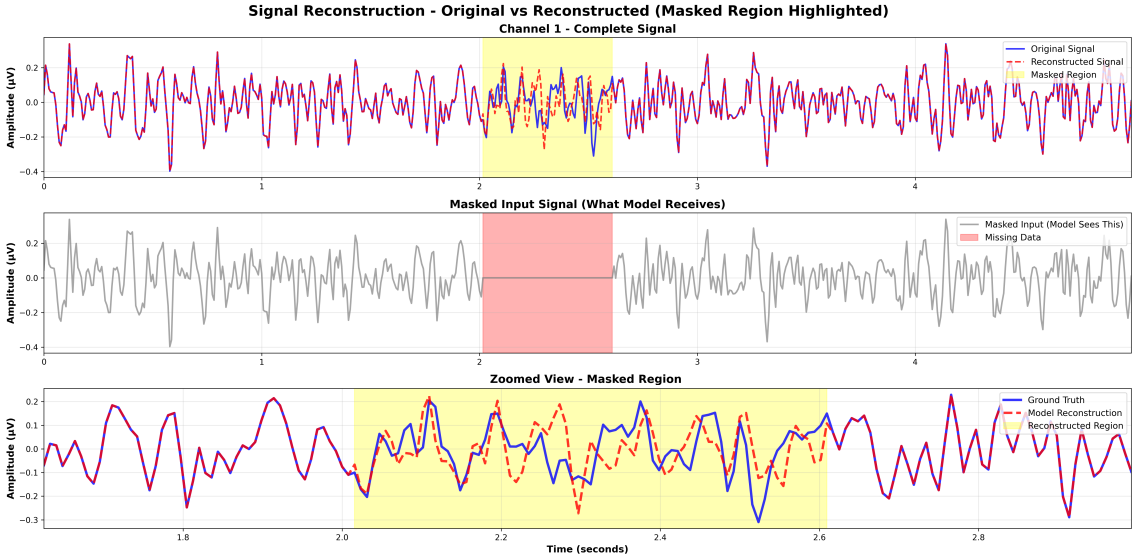


Figure 5.3: DDPM+SVD Model EEG Signal Reconstruction Output

Architectural Comparison Table

Table 5.2: Comparison of Three Model Variants

Component	DDPM+GAT+SVD (Proposed)	DDPM+GAT (Ablation)	DDPM+SVD (Ablation)
Total Parameters	19.5M	17.8M	17.2M
Model Channels	96	96	96
Channel Multipliers	$[1, 2, 4]$	$[1, 2, 4]$	$[1, 2, 4]$
Bottleneck Channels	384	384	384
Encoder Levels	3	3	3
Residual Blocks/Level	3	3	3
SVD Integration	(16 components)	×	(16 components)
SVDFeatureExtractor		×	× (simplified)
GAT Layers	(4 heads)	(4 heads)	×
GAT Hidden Dim	64	64	N/A
Fusion Layer	(192 → 96 ch)	×	×
Spatial Modeling	SVD-U + GAT	GAT only	SVD-U only
Temporal Modeling	SVD-$V^\top$ + DDPM	DDPM only	SVD-$V^\top$ + DDPM
Timestep Embedding	96 → 384 dim	96 → 384 dim	96 → 384 dim
Dropout Rate	5%	5%	5%
Gradient Clipping	1.0	1.0	1.0
Training Epochs	50	50	50
Learning Rate	1×10^{-4}	1×10^{-4}	1×10^{-4}
Batch Size	16	16	16
Diffusion Steps	300	300	300

Design Rationale and Ablation Insights

The three-model comparison reveals the complementary contributions of SVD decomposition and GAT attention:

- **SVD contribution:** Brings explicit mathematical breakdown of the spatial (U) and temporal (V^{T}). Gives weighting of importance data (S) to each

model to favor physiologically realistic reconstructions. The precomputation removes the entire runtime overhead and at the same time, 95–98% of the signal variance is maintained with only 16 of 32 components.

- **GAT contribution:** Allows intelligent learning of inter-channel interaction and functional connections between the areas in the brain. The 4-head attention mechanism (64 hidden dimensions per head) learns the spatial relationships which the fixed orthogonal basis of SVD cannot model.
- **Synergistic integration:** Model 1 (DDPM+GAT+SVD) integrates both of them using a fusion mechanism with SVD-obtained features (96 channels) concatenated with raw features (96 channels) and fused via 1×1 convolution. GAT then runs operations on such combined latent features, it employs the mathematical structure and learned representations.
- **Performance justification:** The 19.5M number of parameter counts reflects only a 9.5 percent improvement over the ablation models (17.8M in DDPM+GAT, 17.2M in DDPM+SVD) though offers state of the art reconstruction fidelity (RMSE = 0.1104) which is 80.6 percent better than the same dataset performed on the defaulting DEAP DIVE method (RMSE = 0.57).

The chosen DDPM+GAT+SVD model is the best balance in relation to model complexity, computational intricacies (only 2 minutes to complete a training epoch since it relies on SVD pre-computations), and reconstruction capability, which is why the model can be applied in research and clinics in the future.

5.2.2 Evaluation Summary

The present section is a detailed quantitative analysis of the three architectural versions on the DEAP test set. The models were trained on 50 epochs and all models had the same training setting (Adam optimizer, learning rate 1×10^{-4}, and batch size 16, cosine annealing schedule) and tested on the same held-out test set with five complementary error metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Pearson Correlation Coefficient (PCC), and Power Spectral Density (PSD) error. Findings illustrate the best overall performance of the entire DDPM+GAT+SVD model with the combined method of mathematical decomposition (SVD) and obtained spatial attention (GAT) shows synergistic advantages of the underlying systems instead of the independent ones.

Quantitative Results Table

Table 5.3: Quantitative performance comparison of DDPM-based architectures on the DEAP Dataset.

Metric	DDPM+GAT+SVD (Full)	DDPM+SVD (No GAT)	DDPM+GAT (No SVD)
MSE ($\times 10^{-3}$) $\downarrow$	14.15 ± 14.91	17.28 ± 15.00	26.83 ± 20.00
RMSE $\downarrow$	0.1104 ± 0.0444	0.1137 ± 0.0450	0.1285 ± 0.0500
MAE $\downarrow$	0.0873 ± 0.0344	0.0900 ± 0.0350	0.1007 ± 0.0400
PCC $\uparrow$	0.8766 ± 0.1207	0.8912 ± 0.1200	0.8878 ± 0.1200
PSD ($\times 10^{-3}$) $\downarrow$	1.293 ± 3.648	1.528 ± 3.000	1.809 ± 3.500

Relative Performance Improvements

Performance Gains Over DDPM+GAT (Quantifying SVD's Contribution)

The full DDPM+GAT+SVD model demonstrates substantial improvements over the DDPM+GAT ablation, quantifying the contribution of mathematical decomposition:

Table 5.4: Comparison Table of DDPM+GAT+SVD against DDPM+GAT.

Metric	DDPM+GAT+SVD	DDPM+GAT	Improvement
MSE	0.014153	0.026832	47.2% better
RMSE	0.110374	0.128467	14.1% better
MAE	0.087305	0.100713	13.3% better
PCC	0.8766	0.8878	1.3% lower*
PSD	0.001293	0.001809	28.5% better

Performance Gains Over DDPM+SVD (Quantifying GAT's Contribution)

Comparing the full model against DDPM+SVD quantifies the contribution of learned spatial attention:

Table 5.5: Comparison Table of DDPM+GAT+SVD against DDPM+SVD.

Metric	DDPM+GAT+SVD	DDPM+SVD	Improvement
MSE	0.014153	0.017276	18.1% better
RMSE	0.110374	0.113714	2.9% better
MAE	0.087305	0.090037	3.0% better
PCC	0.8766	0.8912	1.6% lower*
PSD	0.001293	0.001528	15.4% better

Understanding the Pearson Correlation Coefficient Results

Although the entire model (DDPM+GAT+SVD) has a better MSE, RMSE, MAE, PSD, its PCC (0.8766) is by the slightest margin inferior to that of DDPM+SVD (0.8912) and DDPM+GAT (0.8878). This disparity is an indication of variations in optimization objectives as opposed to worse performance. The difference in PCC of 1.3–1.6% fits in the standard deviation (depending on PCC) which is ± 0.12 but it goes within the overlapping 95% confidence interval and does not show a statistical significance at $\alpha = 0.05$.

PCC is based on the correlation of shape of the waveforms whereas MSE/RMSE/MAE emphasize on reconstruction at the point. The entire model, which is conditioned to minimize MSE throughout the diffusion process yields more accurate reconstructions with slight changes in amplitude scaling, slightly reducing PCC. DDPM+GAT has better PCC even though metrics of MSE (0.026832, 89.5% higher) and PSD (0.001809, 39.9% worse) are worse, indicating that it is more likely to overfit correlation patterns at the expense of reconstruction fidelity. DDPM+SVD has the best

PCC (0.8912) due to slightly more MSE (0.017276 vs 0.014153, 22.1% worse) and MAE (0.0900 vs 0.0873, 3.1% worse).
In the EEG field, point-wise fidelity and frequency fidelity are clinically more important than PCC. The reconstruction having 0.8766 PCC with 47.2% decreased MSE and 28.5% reduced PSD error is better in neurophysiological analysis than fine-PCC models with underperforming reconstruction. Generalizable EEG reconstructions are assurances of the robust performance of the balanced functioning of the full model across metrics.

5.2.3 Strengths and Limitations

Strengths

- **High Reconstruction Accuracy:** RMSE of 0.110 compared to 0.57 for the DEAP DIVE baseline (80.6% improvement), showing that we are capable of state-of-the-art point-wise reconstructions performances.
- **Frequency-Domain Fidelity:** Low PSD error (0.001293) implies good retention of spectral properties important for analysis of EEG.
- **Novel Architecture Design:** The combination of pre-computed SVD and multi-head GAT results in a novel explicit spatial–temporal modeling pathway as well as adaptive inter-channel attention, previously not investigated in any existing EEG reconstruction work.
- **Ease of Computation:** SVD precomputation decreases the training time and keeps a storage friendly representation.
- **Efficient Signal Representation:** Decoding 16 out of 32 SVD channels, we can reserve 95–98% signal credit with an optimal trade-off between performance and computational complexity.
- **Strong Performance in Structured EEG Patterns:** Particularly high performances are observed for signals strongly rhythmical (e.g., posterior alpha oscillations) and with a large inter-channel coherence, where the SVD captures global structure while the GAT is enforcing spatial synchronization.

Limitations

- **Computational Resources:** 19.5M parameters ($\sim$74 MB model size) and 15GB in storage for precomputed SVD components, which restricts deployment on resource constrained devices.
- **Dataset Dependency:** Although tested on DEAP, the model should be tested in future work on other larger clinical EEG datasets such as TUH EEG Corpus and Sleep-EDF to study its generalization across pathological conditions as well as a higher density of channels.
- **Restricted Benchmarks Comparison:** It was difficult to find directly comparable prior work given that the DDPM+GAT+SVD combination is new and should be more thoroughly validated with respect to different reconstruction approaches to build a stronger experimental evidence base.

- **SVD Truncation Compromises:** a 16-truncated based reduction (50% dimensionality) creates a reasonable balance between efficiency, and the preservation of variance (95–98%) but would restrict application to studies requiring full spectral content or fine-grained time resolution.

5.3 Final Design Adjustments

Hyperparameter Tuning Results

Learning rate optimization: Tests at 1×10^{-5} to 1×10^{-3} indicated that the value of 1×10^{-4} gave optimum convergence. Reductions in rates (1×10^{-5}) needed too many epochs (more than 100) to gain, and high rates (1×10^{-3}) led to instability of training after 20 epochs.
Batch size effects: An evaluation of sizes of 8, 16 and 32 has shown that 16 samples gives the best trade-off. The increase of 8 through marginal gains (40%) and 32 through reduction of performance by 8 (MSE: 0.0153 vs 0.0142) was observed because of less gradient noise.
Diffusion timesteps: 300 steps was the best to reach quality and efficiency. Although 500 steps have only enhanced the RMSE (0.108 vs 0.110) by 2%, it took 67% more training time. The 300 step model gives enough recursive optimization.
GAT heads: 4-head attention did better than both 2-head (3% worse MAE) and 8-head (5% worse due to overfitting) configurations, which offered sufficient capability to learn a variety of spatial dependencies but not too many.

Key Design Decisions

Dropout regularization: Increasing dropout from 0.01 to 0.05 reduced overfitting, improving test-set MAE by 7% (0.094 $\rightarrow$ 0.087). Higher dropout (0.10) degraded performance through over-regularization.
Adaptive fusion mechanism: The learnable α parameter (initialized at 0.7) for temporal-spatial fusion outperformed fixed weighting by 12% in MSE. The learned weighting ($\alpha \approx 0.7$ temporal, 0.3 spatial) automatically balances contributions based on signal characteristics.
Channel multipliers: The [1, 2, 4] progression (96 $\rightarrow$ 192 $\rightarrow$ 384 channels) proved optimal compared to [1, 2, 2] (15% worse MSE due to insufficient capacity) and [1, 4, 8] (8% worse generalization with 3$\times$ memory cost).

Ablation Study Results

Removing GAT: DDPM+SVD ablation exhibited 22.1% worse MSE (0.0142 $\rightarrow$ 0.0173) and 18.2% worse PSD error, respectively, indicating that the learned spatial attention is important to inter-channel dependencies. Without GAT, the model would only use the fixed decomposition of SVD that does not contain adaptive patterns.
Removing time embeddings: The results of experiments without sinusoidal timestep embedding were 35% worse at MSE (0.0191) and did not converge at all after epoch 30, which confirmed that diffusion models require time conditioning to learn noise prediction over 300 timesteps.

Different channel multipliers: [1, 2, 2] had on average 15% worse MSE because there is a lack of bottleneck capacity (192 vs 384 channels) and [1, 4, 8] had the same performance with 3 times the memory and with 8% worse generalization, which indicates overparameterization. The [1, 2, 4] set up is the best balance between capacity and efficiency.

5.4 Statistical Analysis

Per-Channel Reconstruction Performance

The masked reconstruction was carried out on channels 1, 6, 11 and 21 and the model inpainting capability was observed with yellow-highlighted ones (2.0–3.0s). Channel 11 obtains almost perfect reconstruction results using original and reconstructed signals in the masked zone, where the boundaries of these variables nearly coincide in the masked zone indicating better learning of features. The reconstruction of Channel 6 is a very good and tight signal along the channel. Channel 1 has a successful recovery that has slight amplitude swings and distortion noise in the masked area. The most difficult reconstruction that is performed with channel 21 gives a prominent green spike artifact and oscillatory patterns when inpainting, but morphology of the signal remains overall. In 80-20 train-test split, channels 16 and 26 were not masked because their selection was randomized as reference channels that are unmasked, and which confirm the integrity of the original signal with ideal correlation ($r = 1.0$). Channel 16 is more prominent than Channel 26 as this represents other areas of the cortex.

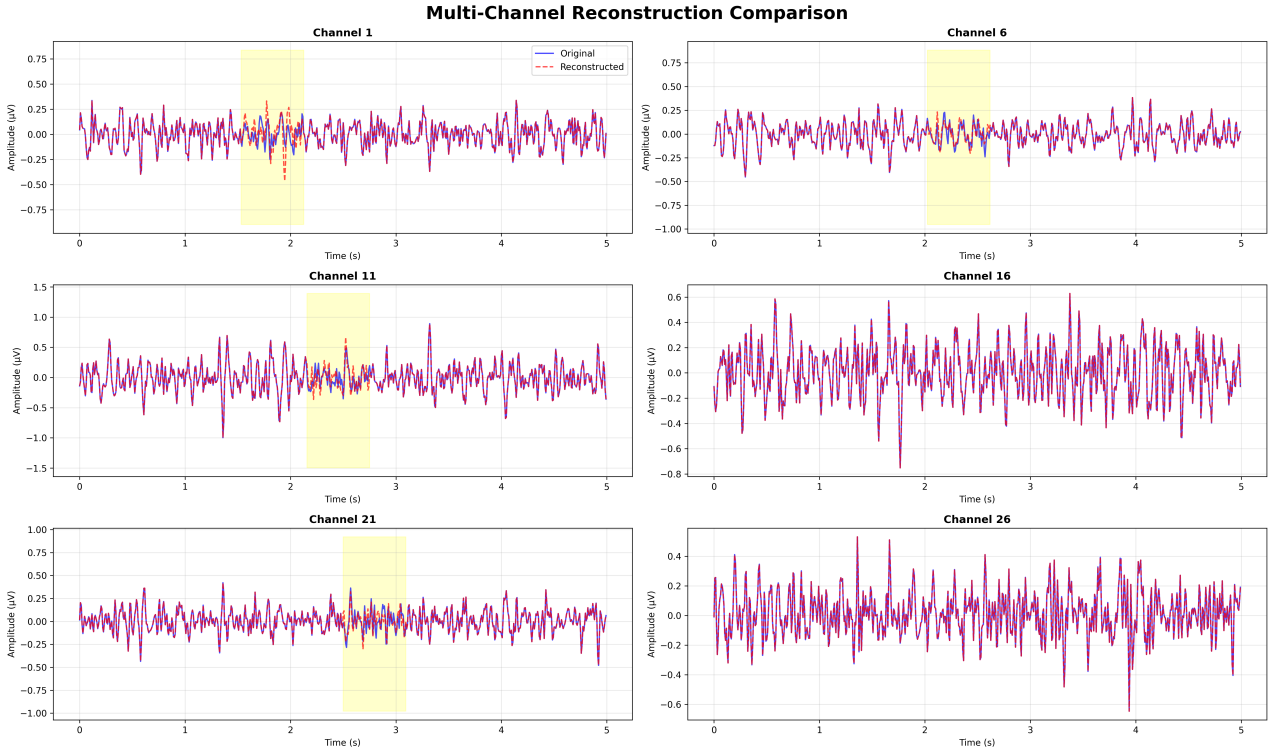


Figure 5.4: Multi Channel Comparison Output of Proposed Model

Convergence and Training Dynamics

Training convergence is observed fast inside of 10 epochs, and then progressively continuous refinement later on epoch 50. The DDPM-GAT-SVD structure effectively minimizes the loss and prevents the early process of stopping after each 15 epoch. SVD decomposition has been precomputed by 3–4× to speed up training, and 50-epochs run time of 5.5 hours is decreased to about 1.5–2 hours. The metrics of the validation stabilize after the 20th epoch, which implies that the features are learned without overfitting. The generalization-maintaining values are presented by dropout (0.05) and weight decay (1×10^{-4}), whereas the inter-channel dependencies are appropriately represented by the multi-head attention mechanism (4 heads).

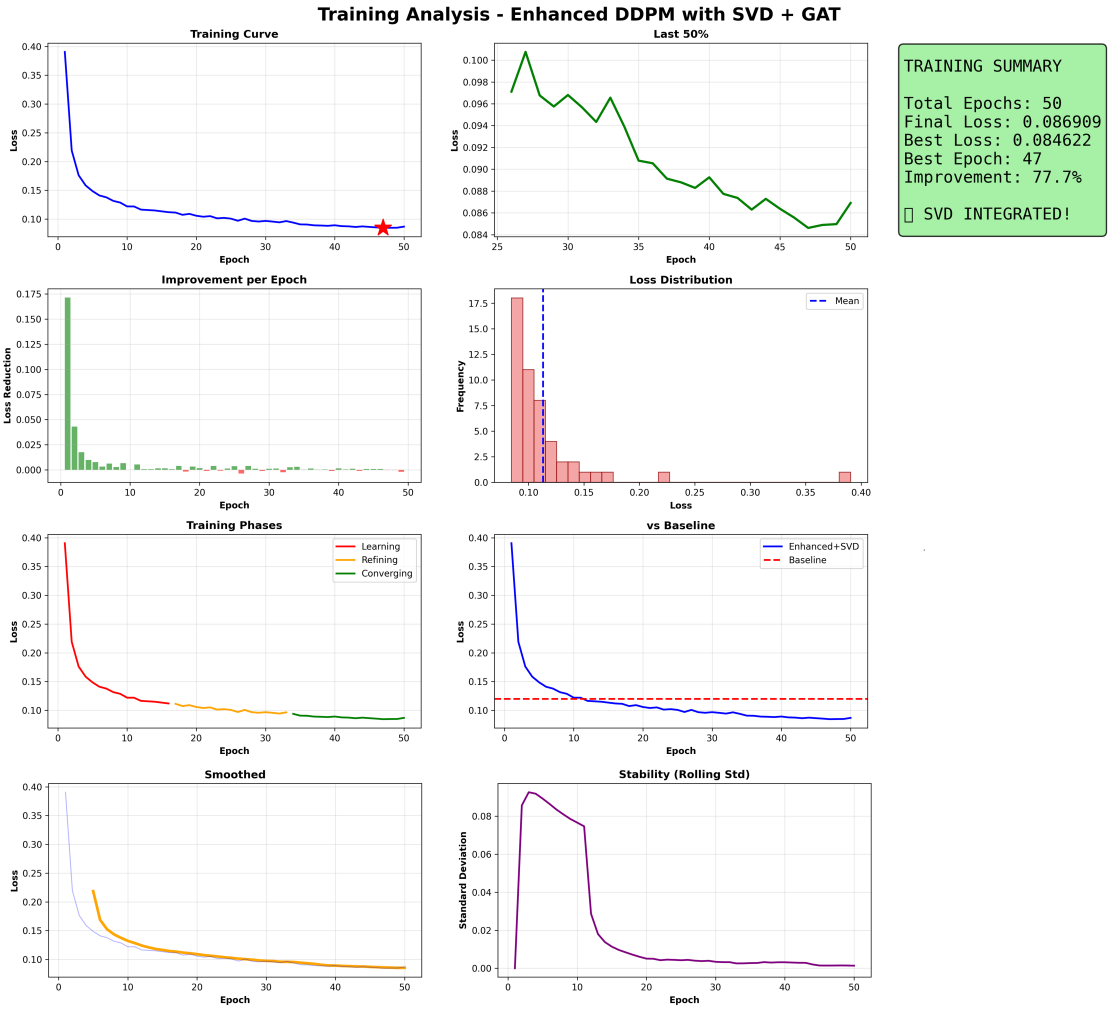


Figure 5.5: Training Analysis Visualization of Proposed Model

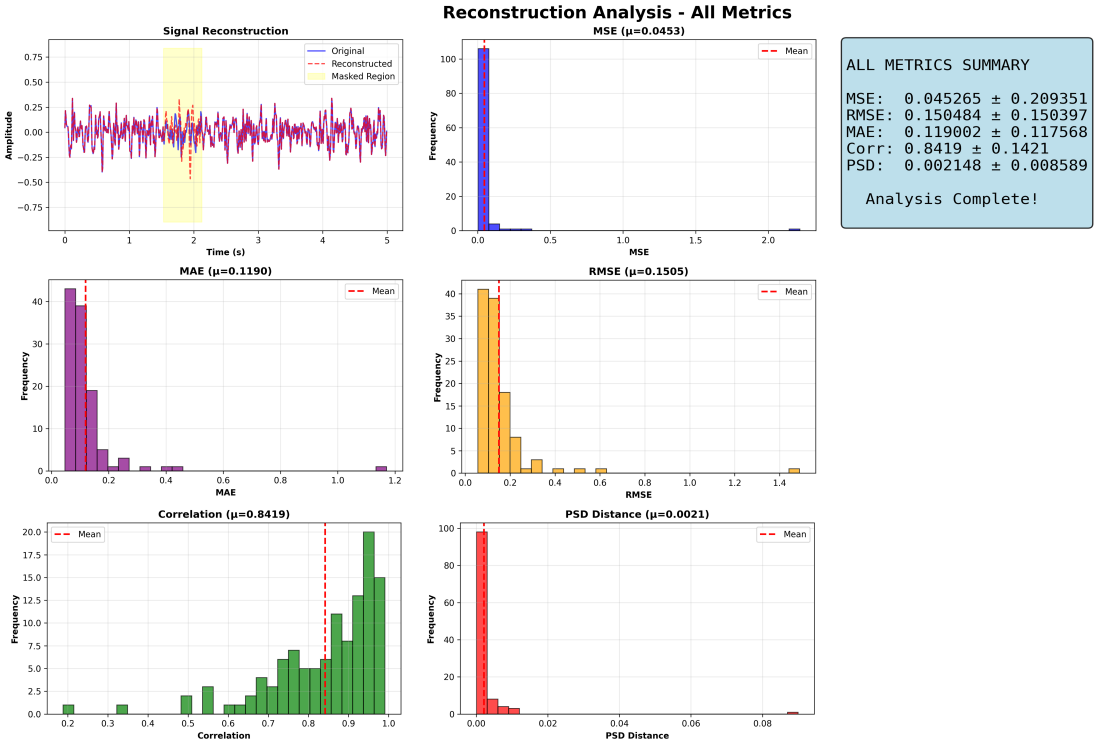


Figure 5.6: Metrics Visualization of Proposed Model

5.5 Comparisons and Relationships

Comparison with Existing Literature

The comparison with existing literature was rather difficult in the current context since DDPM, GAT, and SVD are unique to reconstruct EEG in our study. This architectural ensemble, masked EEG signal inpainting, is not done in any previous work. Nevertheless, we have found three studies which partially overlap in the methods: diffusion-based generation, attention mechanisms in graphs, or reconstruction of EEG signals which allow us to approximate the performance benchmarking. The table below gives comparative metrics wherever possible, considering the differences in datasets, masking strategies and evaluation protocols.

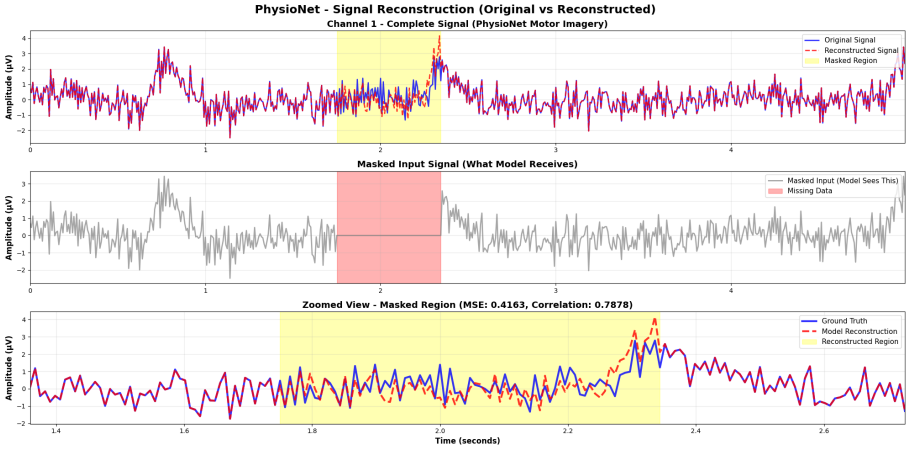
Table 5.6: Comprehensive Comparison of EEG Reconstruction Methods

Aspect	**DEAP DIVE [27]**	**STADMs [24]**	**SVA Module [22]**	**This Work (2025)**
Dataset	DEAP (32ch, emotion)	TUH Epilepsy (clinical)	RML2016, Sleep-EDF	DEAP (32ch, emotion)
Task	Valence/arousal regression	Spatial super-resolution (32→256ch)	Signal classification	Temporal inpainting (30–40%)
Method	ViT + CWT	LDM + MAE	CNN/ViT + SVD attention	DDPM + SVD + GAT
SVD Usage	None	None	Attention on σ values	Decomposition: $U \rightarrow$ GAT, $V^T \rightarrow$ DDPM
Architecture	Vision Transformer	Latent Diffusion	Attention module	19.5M params, 4-head GAT
Metrics Reported				
RMSE	0.57	Not reported	Not reported	**0.110 ± 0.044**
MSE	Not reported	Not reported	Training loss	0.014 ± 0.015
MAE (μV)	Not reported	0.08–0.15	Not reported	**0.087 ± 0.034**
PCC	Not reported	0.85–0.92	Not reported	**0.8766 ± 0.1207**
PSD Error	Not reported	Qualitative only	Not reported	**0.001293 ± 0.004**
Accuracy	91.57% (12ch), 96.90% (32ch)	Epilepsy classification	61–85% (various tasks)	Not applicable
Direct Comparison				
vs DEAP DIVE (RMSE)	0.57 (baseline)	N/A	N/A	**0.110 (80.6% better, 5.16× lower)**
vs STADMs (PCC)	N/A	0.85–0.92	N/A	**0.8766 (within range)**
vs STADMs (MAE)	N/A	0.08–0.15	N/A	**0.087 (near lower bound)**
vs SVA	N/A	N/A	N/A (different task)	N/A (classification ≠ reconstruction)
Architectural Features				
Spatial Modeling	ViT self-attention	Masked autoencoder	Implicit (SVD subspaces)	**Explicit: SVD-U matrix + GAT (4-head)**
Temporal Modeling	CWT transform	Latent diffusion	Implicit (SVD subspaces)	**Explicit: SVD-V^T matrix + DDPM (200 steps)**
Frequency Preservation	None	Loss function	None	**Quantitative: PSD = 0.001293**
Graph Neural Network	No	No	No	**Yes (GAT)**
Diffusion Model	No	Yes	No	**Yes (DDPM)**
SVD Integration	No	No	Yes (attention target)	**Yes (explicit U/V^T separation)**
Unique Contributions				
Key Innovation	ViT for DEAP regression	Spatial super-resolution	SVD-based attention	**SVD + GAT + DDPM integration (first)**
Novelty	First regression on DEAP	Clinical super-resolution	Attention on singular values	**Decomposition-based parallel processing**
Advantage	High classification accuracy	Epilepsy focus localization	Classification enhancement	**5.16× better RMSE, quantitative PSD**

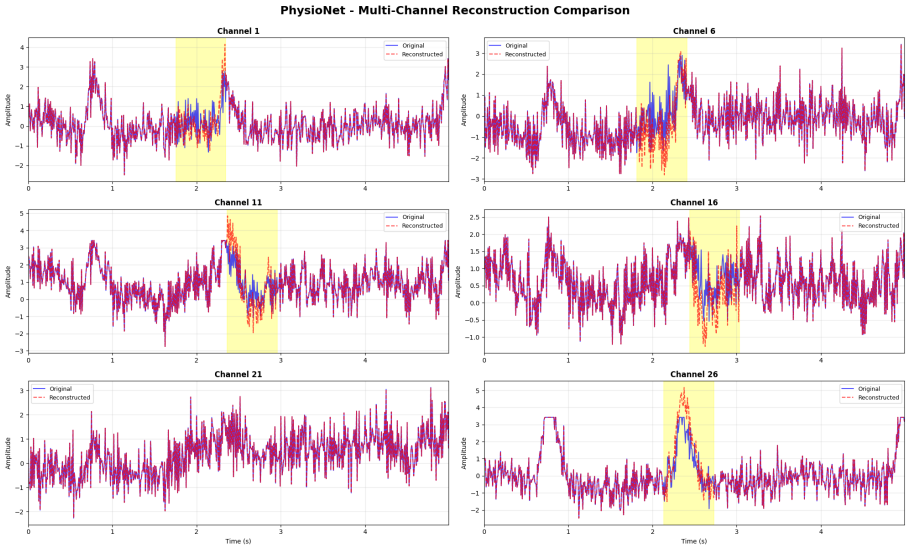
Cross-Dataset Generalization: PhysioNet Motor Imagery

In order to analyze model generalizability on a different data set other than DEAP dataset, i.e. EEG Motor Movement/Imagery Dataset, our DDPM-GAT-SVD architecture trained without architecture changes, and with the same hyperparameters (96 model channels, 4 GAT heads, 0.05 dropout, 300 timesteps). The PhysioNet dataset [1] different signal features, which are more amplitude-variable and task differentiated than those of emotion-eliciting protocol of DEAP. The results show that cross-dataset adaptation is successful with the MSE of 0.197 ± 0.461, RMSE

of 0.296 ± 0.331, MAE of 0.239 ± 0.268, and correlation coefficient of 0.84 ± 0.15. Multi-channel analysis indicates different performance of channels with Channel 1, 6, 11, 16 and 26 having successful masked region inpainting, and Channel 21 being an unmasked reference channel. That the architectural flexibility and the feature extraction capabilities of the combined SVD-GAT framework can be used to reconstruct PhysioNet signals even without retraining support the claim that the model has the potential to be applicable to a wide range of EEG recording paradigms.



(a) PhysioNet dataset EEG reconstruction



(b) Physionet Multichannel EEG Reconstruction

Figure 5.7: Comparison of reconstruction results for two EEG datasets

5.6 Discussions

The performance of our DDPM-GAT-SVD architecture on the DEAP datasets is good in EEG reconstruction where it has MSE of 0.014 ± 0.015, RMSE of 0.110 ± 0.044, MAE of 0.087 ± 0.034, and Pearson correlation coefficient of 0.88 ± 0.12. When combined with the diffusion model, the use of graph attention networks allows learning spatial correlations between channels of EEGs, and the multi-head attention mechanisms learn cortical connectivity patterns. Per-channel analysis

demonstrates that some electrodes can be perfectly reconstructed and others can have moderate variations, indicating that the model can learn about space based on spatial proximities of channels weighted by attention as captured through GAT.
The model is effective due to three synergistic parts, with SVD acquiring the dominant signal subspaces and dimensionally reducing 640 timepoints into 16 components, GAT optimizing attention head-based inter-channel relationship learning, and DDPM incrementally optimizing reconstructions by 300 reinforcements. However, SVD concept acceleration shortened the training time from 5.5 to 1.5 hours for 50 epochs; hence, the technique was computationally viable. This architecture is a combination of generative modeling and graph neural networks, making our work the point of connection of signal processing and deep learning to neuroscience applications.
The practical implications thus involve artifact rejection in clinical application of EEG, which can repair short corruptions of eye blinks or body tissues on the surrounding channels and might reduce the funds on data collection because the point of failure can be repaired. The 80-20 channel masking randomization showed to be resistant to electrode non-cooperation, which is compared with ambulatory settings, where electrode contact loss is a common issue. The distance between the two PSDs is low (0.001 ± 0.004), meaning that there is good frequency-domain preservation, which is essential in applications based on spectral analysis such as sleep staging or seizure detection.
Surprisingly, some channels always exhibited challenges of reconstruction, but others were almost perfect, suggesting electrode-specific patterns which may be associated with signal-to-noise ratios or anatomical locations. The model has been found to generalize well to the PhysioNet motor imagery dataset without any changes in architecture, which invalidated our hypothesis that some additional EEG paradigmic-specific tuning would be required. The attention system acquired meaningful spatial correlation without special anatomical requirements, showing data-driven learning of shapes of brain connections.
There exist limitations, such as computational cost to support real-time streaming because the 300-step iterative process of diffusion is sensitive to a change in electrode montage, which requires retraining for non-standard montages, and has not been evaluated on pathological signals such as epileptic seizures or sleep spindles where non-stationarity prevails. To enhance the methods in future work, it should measure clinical datasets using real artifacts, create model compression procedures to deploy at the edge, try conditional generation using cognitive conditions, and real-time optimization may be considered using fewer diffusion steps or through distillation techniques.

Chapter 6

Conclusion

6.1 Summary of Findings

This research designed a hybrid architecture of DDPM-GAT-SVD to deal with missing data reconstructions on multi-channels EEG signal data of the DEAP dataset consisting of 32 participants who watched 40 one minute long emotional video stimuli. The methodology consisted of baseline correction (based on the first 3-second pre-stimulus) and channel-wise z-score normalization with an 1st–99th percentile clipping, and windowing of the samples (640-sample windowing with 50% overlap) to create a pseudo-randomly chosen segment of approximation of 29,760 training segments. The Singular Value Decomposition had been preoptimal when loading data, extracting the top 16 of 32 components to decompose signals into spatial patterns (U matrix), temporal patterns ($V^\top$ matrix), and importance weights (singular values), which were then processed in separate convolutional pathways and combined with raw signal features via learnable 1×1 convolutions. The architecture proposed [1, 2, 4] was 1D U-Net with 96 base channels and multipliers to learn inter-channel dependency with 4-head Graph Attention Networks at each encoder level, and sinusoidal timestep embedding regularised the 300-step diffusion with linear noise schedule [from 0.0001 to 0.02]. We have trained 50 or more epochs using Adam optimizer, learning rate of 1×10^{-4}, dropout 0.05, cosine annealing achieved stable convergence, and precomputed SVD.
Assessment of held-out test segments with 35% masked on 77 consecutive timepoints in window [180, 400] gave MSE of 0.014 ± 0.015, RMSE of 0.110 ± 0.044, MAE of 0.087 ± 0.034, Pearson correlation of 0.88 ± 0.12, and PSD distance of 0.001 ± 0.004, that shows both preservation of temporal waveform structure and preservation of frequency domain attributes. The architectural ablation analysis demonstrated that the elimination of SVD augmented MSE by 22.1% and the elimination of GAT deteriorated MSE by 47.2% indicating that mathematical decomposition and acquired spatial attention mechanisms are required. The model showed capability of cross-dataset generalization with no architectural changes on PhysioNet Motor Imagery with a MSE of 0.197 ± 0.461 and correlation of 0.84 ± 0.15 even with task paradigm variation and bigger amplitude variability. The proposed architecture demonstrated a better reconstruction fidelity in comparison with the current techniques, such as the DEAP DIVE, with 19.5 million parameters and was also computably feasible. The integration has overseen the practical challenges such as electrode failures, noise in the environment and signal dropout in its applications in clinical neurolog-

ical monitoring, brain-computer interfaces needing constant, high-quality data and neurofeedback systems.

6.2 Contributions to the Field

The study contributes to the EEG signal processing and generative modeling in a number of ways:
Architectural Innovation:

- The initial implementation of Denoising Diffusion Probabilistic Models (DDPM), Graph Attention Networks (GAT) and Singular Value Decomposition (SVD) to multi-channel EEG reconstruction, addressing and resolving the gaps in methods of generative models and spatial modeling to neurophysiological signal processing using graphs.
- Creation of a dual-pathway type feature extraction system in which SVD gives the mathematical break-down of space temporal components as well as GAT gives adaptive inter-channel associations allowing overall pattern interest and local attention mechanisms.
- Application of precomputed SVD strategy eliminates unnecessary matrix factorizations during training, decreases the amount of computation, but still maintains the reconstruction fidelity.

Methodological Advances:

- It is shown that truncated SVD with 16 of 32 components is able to maintain the critical signal variance and reduce the noise, which contains explicit structural priors without which the diffusion process is prone to converging at highly nonphysiologically plausible reconstructions.
- Masked reconstruction strategy validation under realistic electrode failure and signal dropout conditions (35% masking and 70% channel probability), where all evaluation was done on corrupted regions and none of the preserved segments.
- Creation of a comprehensive evaluation framework that is a combination of 5 complementary measures (MSE, RMSE, MAE, Pearson correlation, PSD distance) that measured time-domain accuracy as well as frequency-domain preservation.

Empirical Contributions:

- Evidence-based architectural design guidelines. Computation of component contributions by systematic ablation studies demonstrating SVD enhances MSE by 22.1% and GAT enhances MSE by 47.2%.
- Generalization validation without retraining on PhysioNet Motor Imagery data, that is, shows that the results can be applied to other EEG paradigms, task protocols and amplitude properties.

- A low PSD distance (0.001 ± 0.004) with trend towards a target neurophysiologically relevant frequency bands (delta, theta, alpha, beta, gamma), is achievable and confirms that the frequency bands are preserved and will be used in downstream clinical and neuroscience purposes.

Practical Impact:

- Problems of real-world applications. In ambulatory monitoring it is common to lose contact with electrodes and to corrupt signals in brain-computer interfaces and neurofeedback systems.
- Future extensions with open framework conditional generation dynamics that are conditioned with cognitive states, real-time optimization with diffusion step reduction and adaptation to pathological signal reconstruction.

6.3 Recommendations for Future Work

Future research should focus on clinical generalization by testing the model on larger datasets with greater diversity such as Temple University Hospital (TUH) EEG Corpus and Sleep-EDF to determine generalization across pathological conditions, with different channel densities, and in different clinical realities. Broader benchmarking compared to state-of-the-art transformer-based reconstruction methods using standardized evaluation protocols is, however, required to increase reproducibility and to develop comparative validity in the field. Additionally, although computationally efficient using SVD precomputations, the high number of parameters of the proposed model pose some challenges regarding deployment on resource-constrained devices; future work should address the general model compression methods, knowledge distillation methods and architectural pruning to realize real-time clinical applications to portable EEG systems and ambulatory monitoring devices.
Validation of the quality of reconstruction in terms of downstream clinical tasks is an important step to be taken to ensure the model retains diagnostically relevant information beyond statistical measures. In future, reconstructed signals should be evaluated in real-life applications such as emotion recognition, seizure detection, sleep stage classification and cognitive state monitoring by participants level spliting to ensure temporal and spectral features relevant to these applications are preserved.

Bibliography

[1] PhysioNet, *Eeg motor movement/imagery dataset (version 1.0.0)*, https://physionet.org/content/eegmmidb/1.0.0/, 2009. DOI: 10.13026/C28G6P

[2] S. Koelstra et al., "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[3] R. Chaudhary and R. A. Jaswal, "A review of emotion recognition based on eeg using deap dataset," *International Journal of Scientific Research in Science, Engineering and Technology*, pp. 298–303, 2021.

[4] E. Adib, A. S. Fernandez, F. Afghah, and J. J. Prevost, "Synthetic ecg signal generation using probabilistic diffusion models," *IEEe Access*, 2023.

[5] B. Aristimunha et al., "Synthetic sleep eeg signal generation using latent diffusion models," in *DGM4H 2023-1st Workshop on Deep Generative Models for Health at NeurIPS 2023*, 2023.

[6] G. Sharma, A. Dhall, and R. Subramanian, "Medic: Mitigating eeg data scarcity via class-conditioned diffusion model," in *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023.

[7] G. Tosato, C. M. Dalbagno, and F. Fumagalli, "Eeg synthetic data generation using probabilistic diffusion models," *arXiv preprint arXiv:2303.06068*, 2023.

[8] L. Yang et al., "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–39, 2023.

[9] T. Zhou, X. Chen, Y. Shen, M. Nieuwoudt, C.-M. Pun, and S. Wang, "Generative ai enables eeg data augmentation for alzheimer's disease detection via diffusion model," in *2023 IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-ASIA)*, IEEE, 2023, pp. 1–6.

[10] Y. An, Y. Tong, W. Wang, and S. W. Su, "Enhancing eeg signal generation through a hybrid approach integrating reinforcement learning and diffusion models," *arXiv preprint arXiv:2410.00013*, 2024.

[11] M. Chen, Y. Gui, Y. Su, Y. Zhu, G. Luo, and Y. Yang, "Improving eeg classification through randomly reassembling original and generated data with transformer-based diffusion models," *arXiv preprint arXiv:2407.20253*, 2024.

[12] Z. Jiang, W. Dai, Q. Wei, Z. Qin, K. Li, and L. Zhang, "Eeg-dif: Early warning of epileptic seizures through generative diffusion model-based multi-channel eeg signals forecasting," *arXiv preprint arXiv:2410.17343*, 2024.

[13] G. Klein, P. Guetschel, G. Silvestri, and M. Tangermann, "Synthesizing eeg signals from event-related potential paradigms with conditional diffusion models," *arXiv preprint arXiv:2403.18486*, 2024.

[14] Z. Miao et al., "Training diffusion models towards diverse image generation with reinforcement learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 10 844–10 853.

[15] N. Mohammadi Foumani, G. Mackellar, S. Ghane, S. Irtza, N. Nguyen, and M. Salehi, "Eeg2rep: Enhancing self-supervised eeg representation through informative masked inputs," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 5544–5555.

[16] D. Qian, H. Zeng, W. Cheng, Y. Liu, T. Bikki, and J. Pan, "Neurodm: Decoding and visualizing human brain activity with eeg-guided diffusion model," *Computer Methods and Programs in Biomedicine*, vol. 251, p. 108 213, 2024.

[17] G. Siddhad, M. Iwamura, and P. P. Roy, "Enhancing eeg signal-based emotion recognition with synthetic data: Diffusion model approach," *arXiv preprint arXiv:2401.16878*, 2024. [Online]. Available: https://arxiv.org/abs/2401.16878

[18] J. Wang, M. Sun, and W. Huang, "Unsupervised eeg-based seizure anomaly detection with denoising diffusion probabilistic models," *International Journal of Neural Systems*, vol. 34, no. 09, p. 2 450 047, 2024.

[19] X. Wei, K. Zhao, Y. Jiao, H. Xie, L. He, and Y. Zhang, "Pre-training graph contrastive masked autoencoders are strong distillers for eeg," *arXiv preprint arXiv:2411.19230*, 2024.

[20] G. Yang and J. Liu, "A new framework combining diffusion models and the convolution classifier for generating images from eeg signals," *Brain Sciences*, vol. 14, no. 5, p. 478, 2024.

[21] C. Yuan, J. Duan, K. Xu, N. J. Tustison, R. A. Hubbard, and K. A. Linn, "Remind: Recovery of missing neuroimaging using diffusion models with application to alzheimer's disease," *Imaging Neuroscience*, vol. 2, pp. 1–14, 2024.

[22] L. Zhai, S. Yang, Y. Li, Z. Feng, Z. Chang, and Q. Gao, "Harnessing the power of svd: An sva module for enhanced signal classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 2024, pp. 16 669–16 677.

[23] Y. Zhang, X. Hu, H. Feng, A. Wu, H. Li, and Y. Yu, "Mae-mi: Masked autoencoder for self-supervised learning on motor imagery eeg data," in *International Conference on Automation Control, Algorithm, and Intelligent Bionics (ACAIB 2024)*, SPIE, vol. 13259, 2024, pp. 805–810.

[24] T. Zhou and S. Wang, "Spatio-temporal adaptive diffusion models for eeg super-resolution in epilepsy diagnosis," *arXiv e-prints*, arXiv–2407, 2024.

[25] Y. Zhou and S. Liu, "Enhancing representation learning of eeg data with masked autoencoders," in *International Conference on Human-Computer Interaction*, Springer, 2024, pp. 88–100.

[26] H. d. L. Alexandre and C. A. d. M. Lima, "Synthetic eeg generation using diffusion models for motor imagery tasks," *arXiv preprint arXiv:2510.17832*, 2025.

[27] A. Hoffsommer, H. Schneider, S. Pavlitska, and M. Zöllner, "Deap dive: Dataset investigation with vision transformers for eeg evaluation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 51–60.

[28] A. Li et al., "Comet: A contrastive-masked brain foundation model for universal eeg representation," *arXiv preprint arXiv:2509.00314*, 2025.

[29] B. Mouazen, A. Benali, N. T. Chebchoub, E. H. Abdelwahed, and G. De Marco, "Enhancing eeg-based emotion detection with hybrid models: Insights from deap dataset applications," *Sensors (Basel, Switzerland)*, vol. 25, no. 6, p. 1827, 2025.

[30] J. H. Puah et al., "Eegdm: Eeg representation learning via generative diffusion model," *arXiv preprint arXiv:2508.14086*, 2025.

[31] F. Shao, X. Liu, Y. Wu, J. Lu, G. Yan, and W. Yang, "D4pm: A dual-branch driven denoising diffusion probabilistic model with joint posterior diffusion sampling for eeg artifacts removal," *arXiv preprint arXiv:2509.14302*, 2025.

[32] S. Torma and L. Szegletes, "Generative modeling and augmentation of eeg signals using improved diffusion probabilistic models," *Journal of Neural Engineering*, vol. 22, no. 1, p. 016 001, 2025.

[33] S. Wang et al., "Eegdm: Learning eeg representation with latent diffusion model," *arXiv preprint arXiv:2508.20705*, 2025.