2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385659

LightK-DSGCN: Depression Detection in EEGs with Lightweight Kalman Filter-aided Dual-Stream Graph Convolutional Networks

Yanshen Sun
Department of Computer Science
Virginia Tech
Falls Church, VA, USA
yansh93@vt.edu

Jianger Yu
Department of Computer Science
Virginia Tech
Falls Church, VA, USA
jianger@vt.edu

Chang-Tien Lu
Department of Computer Science
Virginia Tech
Falls Church, VA, USA
clu@vt.edu

***Abstract*—Automated detection of depression using Electroencephalogram (EEG) signals is crucial for advanced disease treatment. However, existing EEG detection models still face challenges. 1) EEG signals are susceptible to noise interference and exhibit a high level of randomness. 2) Manual denoising and feature selection potentially introduce human bias. 3) The integrated message propagation across both spatial and temporal domains is not fully explored. Therefore, this paper proposes LightK-DSGCN, Enhancing Depression Detection with Lightweight Kalman Filter-aided Dual-Stream Graph Convolutional Networks, a novel framework for identifying characteristics EEG patterns of depression patients. LightK-DSGCN leverages dual-stream graph neural networks to simultaneously explore spatiotemporal features, effectively capturing the distinctive patterns exhibited by depression patients. Firstly, the EEG signals are decomposed into temporal and spatial components at each time point. Then, the temporal features are embedded using a dilation temporal convolutional network, while the spatial features are obtained through a graph convolutional network. Moreover, a lightweight Kalman filter combined with recurrent neural networks is proposed to denoise and align the spatiotemporal features, enabling the extraction of detailed information from multiple perspectives. Experimental results on two real-world datasets demonstrate the superiority of our LightK-DSGCN over state-of-the-art methods in detecting depression using EEG signals. LightK-DSGCN provides a promising approach for automated depression detection in clinical practice. The code can be found here.**

***Index Terms*—depression detection, EEG, lightweight Kalman Filter, spatiotemporal feature, dual-stream GCN**

I. Introduction

Depression significantly impairs well-being, with over 350 million affected individuals worldwide in 2023 and an annual mortality rate exceeding one million lives, as the World Health Organization (WHO) reported [1]. This alarming data emphasizes the urgent need for effective depression detection methods due to its high prevalence and substantial underidentification. Among various neurobiological techniques, EEG stands out as an affordable and practical biomarker for investigating the underlying mechanisms of depression, thanks to its advantages of high temporal resolution, non-invasiveness, portability, and affordability [2]. Generally, the detection of depression using EEG signals aims to unveil distinctive patterns in individuals with depression. Nonetheless, this endeavor is fraught with challenges, primarily arising from the EEG signal's susceptibility to noise interference, the signal patterns' complexity and non-linearity, and the mutual dependencies among signal channels. Research on depression detection in EEG signals is still in its early stages and demands further investigation, leveraging computational intelligence, optimization techniques, high-performance computing, and other innovative approaches [3].

Traditional depression detection models primarily concentrate on capturing temporal dependencies. Various techniques, including ARIMA [4], DFT [5], SVR [6], Wavelet Transform [7], and deep learning-based approaches such as Convolutional Neural Network (CNN) [8], and Long Short-Term Memory (LSTM) [9], have been extensively utilized in modeling EEG series data. However, none of these methods can handle the intrinsic geographic dependencies in EEG data, even though some researchers attempted to tackle this issue through CNN [10], [11]. To address this challenge, recent approaches have turned to graph neural networks (GNNs) [12], [13]. However, these methods often aggregate temporal features as the input for GNNs, overlooking the importance of considering the spatiotemporal co-evolution of EEG patterns.

Presently, dynamic GNN modeling techniques have garnered considerable attention in various tasks, including node classification, link prediction, and predicting failure paths in dynamic graphs [14], [15]. Although various dynamic GNN models have been proposed, most are not explicitly tailored to handle the substantial variances that can emerge in EEG signals for depression detection. Detecting depression in EEG signals faces the following challenges: **(1) Existing trainable EEG denoising neural networks are limited in accurately identifying valuable signal patterns.** It is still an open question of how to effectively replace the complex and time-consuming filtering procedures with trainable deep learning modules. **(2) Previous works rely heavily on manual feature**

extraction to handle the noisy nature of EEG signals. However, manual denoising could be limited by human cognitive biases and potentially result in the removal of crucial features. **(3) Current methods for analyzing EEG signals fail to capture the spatiotemporal dependencies adequately.** Most of the existing methods process spatial and temporal relations separately while neglecting joint spatiotemporal message passing.

To tackle these challenges, we present LightK-DSGCN, a novel spatiotemporal graph neural network for depression detection in EEG signals. Our approach leverages trainable denoising modules to conduct end-to-end depression signal classification results. By integrating denoised spatiotemporal information, our model exploits normal patterns while minimizing the influence of noise. The main contributions of this paper can be summarized as follows.

- **Designing a novel spatiotemporal model for depression EEG signal detection** We introduce LightK-DSGCN, an innovative model for depression detection in EEG signals that seamlessly incorporates temporal-spatial networks to capture comprehensive patterns. LightK-DSGCN combines spatial and temporal features while preserving the atomicity of training samples. Additionally, the model leverages diverse neural network architectures and denoising techniques to encode smooth EEG series effectively.
- **Proposing an end-to-end deep learning architecture with multiple novel trainable denoising modules** Instead of complex pre-filtering, LightK-DSGCN identifies depression samples from normalized EEG signals. Specifically, two trainable denoising modules are incorporated in LightK-DSGCN. The dilation temporal convolution module smoothens noisy data along the temporal dimension and the lightweight Kalman filter rectifies high-noise features while retaining low-noise features, allowing for seamless spatial and temporal information integration.
- **Conducting comprehensive experiments to prove the effectiveness of LightK-DSGCN** Experiments were performed on two real-world EEG depression datasets to evaluate the effectiveness and efficiency of LightK-DSGCN. We employed a ten-cross-validation method and multiple metrics to compare the performance of LightK-DSGCN against the performances of the other depression detection models. The experimental results demonstrate that our LightK-DSGCN outperforms state-of-the-art techniques across multiple evaluation metrics.

II. Related work

A. *Time-frequency analysis method*

Previous methods typically involved extracting temporal and spectral features from EEG signals using feature engineering techniques, followed by using machine learning models to classify individuals into depressed and healthy categories [16]–[18]. For instance, Movahed et al. [19] extracted power features from four EEG frequency bands and four nonlinear features, achieving a 90% accuracy using a combination of nonlinear features with classifiers like K-nearest neighbor (KNN), linear discriminant analysis, and logistic regression (LR). Similarly, Spyrou et al. [20] estimated EEG data using orthogonal discrete wavelet transform, employing random forest, decision tree, multilayer perceptron (MLP), and Support Vector Machine (SVM) classifiers with recognition accuracies ranging from 92.42% to 95.45% [21], [22].

B. *Deep learning method*

Deep learning approaches have emerged as powerful tools for depression detection [23]–[25]. Li et al. [26] proposed a mild depression recognition method based on convolutional neural networks (CNN), emphasizing the significance of spatiotemporal features in EEG signals. Sharma et al. [27] and Song et al. [28], combined CNN with LSTM models. Liu et al. [29] utilized CNN and gated recurrent unit (GRU) to extract EEG sequence features for depression recognition. Other studies have focused on constructing brain functional networks based on EEG signals for depression classification [30]–[32]. Chen et al. [12] developed a GCN with attention mechanisms for depression detection. Wu et al. [33] introduced a deep learning method combining a spatiotemporal graph convolution network (ST-GCN) with a depression-related functional connectivity network. Zhu et al. [13] calculated coherence-based functional connectivity features for depression classification.

Existing approaches have underutilized the potential of graph neural networks in analyzing dynamic spatiotemporal features. In this paper, we propose a novel method called LightK-DSGCN. By leveraging dynamic graph neural networks, LightK-DSGCN effectively explores spatiotemporal features, enabling the identification of patterns indicating depression. Through the integration of denoised spatial and temporal information, LightK-DSGCN restores normal patterns while reducing the impact of noises.

III. Preliminary (Lightweight Kalman filter)

The classic Kalman filter is widely employed in regression and deep learning models, thanks to its capability to combine two estimates by minimizing their covariance. Inspired by [34], we leverage the principal concept of the Kalman filter for aligning the outputs of the spatial and temporal analysis modules. The idea of the lightweight Kalman filter is described as follows. For a value x, consider that there are two independent estimations x_1 and x_2, and x is a weighted combination of x_1 and x_2, described as (1).

$$\hat{x} = w_1 x_1 + w_2 x_2 \tag{1}$$

where w_1 and w_2 are weights of x_1 and x_2, $w_1 + w_2 = 1$. The expectation of $\hat{x}$, $E(\hat{x}) = w_1 E(x_1) + w_2 E(x_2)$. As $E(x_1)$ and $E(x_2)$ are independent, $E([x_1 - E(x_1)][x_2 - E(x_2)]) = 0$. σ^2 can then be written as (2).

$$\begin{aligned} \sigma^2 &= E([\hat{x} - E(\hat{x})]^2) \\ &= w_1^2 E([x_1 - E(x_1)]^2) + w_2^2 E([x_2 - E(x_2)]^2) \\ &= w_1 \sigma_1^2 + w_2 \sigma_2^2 \end{aligned} \tag{2}$$

Let $w_2 = w$ and $w_1 = 1 - w$. To minimize σ^2, let the differential between σ^2 and w be zero:

$$\frac{d\sigma^2}{dw} = -2(1-w)\sigma_1^2 + 2w\sigma_2^2 = 0 \tag{3}$$

The analytical solution of w is:

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{4}$$

$\hat{x}$ and σ^2 are:

$$\hat{x} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2}, \sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{5}$$

Based on this mechanism, our LightK-DSGCN model considers two estimations, x_1 and x_2, generated by two different models. From Equation 5, it can be inferred that the merged estimation $\hat{x}$ is the weighted sum of x_1 and x_2, where the weights are $\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ and $\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$ respectively.

IV. LightK-DSGCN Model

Our LightK-DSGCN framework comprises five modules: feature acquisition (F-acquisition), spatial feature stream (SF-stream), temporal feature stream (TF-stream), Lightweight Kalman filter alignment (LKF-alignment), and detection module, as shown in Fig. 1.

At each timestamp t, the SF-stream module takes the adjacency matrix and time sequence data as input, while the TF-stream module processes the preprocessed time sequence data. The SF-stream module uses a GCN block to encode spatial-dependency features, followed by an LSTM layer. Similarly, the TF-stream module encodes multi-scale temporal features using a dilation TCN block and an LSTM layer. In the LKF-alignment module, the outputs from the SF-stream and TF-stream modules are merged using the lightweight Kalman filter and LSTM layers, and these merged features are then passed to the next timestamp $t+1$. At the final timestamp T, the outputs from the LKF-alignment module are fed into the detection module, which generates the final detection results.

A. Feature acquisition

This module extracts distinctive spatiotemporal characteristics from EEG signals captured by each electrode. The extracted temporal features are used in the TF-stream module, while the EEG networks are utilized in the SF-stream module to obtain spatial features of the EEG signal.

*1) **Temporal feature extraction**:* EEG signals are always prone to contamination and artifacts due to environmental and physiological factors. Therefore, it needs to preprocess the raw EEG data to obtain clean and artifact-free signals for further research, including re-referencing, filtering, interpolating bad channels, and removing EEG artifacts [35]. After this preprocessing pipeline, we successfully eliminate unwanted components, such as eye blinks, muscle artifacts, and environmental interference. The process concludes with the extraction of more refined data, ready for subsequent feature extraction.

Temporal feature extraction aims to construct a feature matrix for each participant. Assuming there are N electrodes on the EEG cap, each participant can provide N EEG signal sequences. We divided each signal series into T segments of length l. According to the predetermined sampling frequency, each segment comprises F sampled values of the EEG signal, thereby encompassing F dimensional features in each segment. Consequently, the EEG matrix attributed to each participant is represented as $\mathcal{X} \in \mathbb{R}^{N \times T \times F}$, wherein N denotes the count of electrodes affixed to the EEG cap, T signifies the quantity of segments in an EEG signal sequence, and F represents the number of EEG sampling values contained within each piece.

*2) **EEG network construction**:* The EEG network is defined as an undirected graph, denoted as $G = (V, E, \hat{A})$. The set of nodes V consists of N nodes, each defined by the fixed position of an electrode. The edges connecting the nodes are captured in E, and the adjacency matrix representation of these edges is denoted as $\hat{A} \in \mathbb{R}^{N \times N}$ which is obtained by sparsifying the initial adjacency matrix $A \in \mathbb{R}^{N \times N}$.

We first constructed initial adjacency matrix $A = \{(a_{i,j})\}, i = 0..N-1, j = 0..N-1$, where $a_{i,j}$ is defined as the Euclidean distance between nodes i and node j. This is a dense, weighted, and fully connected graph, where edge weights indicate the degree of node proximity. Based on domain knowledge, although each electrode exhibits a certain level of adjacency with all electrodes, most electrodes are highly adjacent to a few and weakly adjacent to others. This property inspires us to perform sparsification on the initial adjacency matrix A to get our needed EEG network $G = (V, E, \hat{A})$.

Sparsification techniques retain connections to the nearest neighbors while breaking connections to less similar nodes. This approach effectively leverages local information to counter the low spatial resolution in EEG signals, providing a more comprehensive and accurate representation of electrode pair correlations. Moreover, it reduces the impact of noise and outliers. So, adjacency matrix $\hat{A}$ retains only the connections between each electrode and its K nearest neighboring electrodes. To simplify the process, we set the edge weights between each electrode and its K nearest neighbors to 1, and the weights for the remaining edges to 0. After experimentation, we determined k=4 in this study.

B. SF-stream module

Once the EEG networks are constructed, they are fed to the SF-stream module, which integrates GCN and LSTM. Unlike the traditional GCN approach that performs graph convolution in the frequency domain using the Laplacian matrix and approximates the aggregation of "k-hop neighbors" [36], the SF-stream module employs a straightforward version of GCN without any approximation.

The method can be seen as a linear combination of weights assigned to the $1, 2, ..., k$-nearest neighbors of each node. This allows for dynamic adjustment of the neighbor effects through the coefficients. To compute the weights of the kth nearest

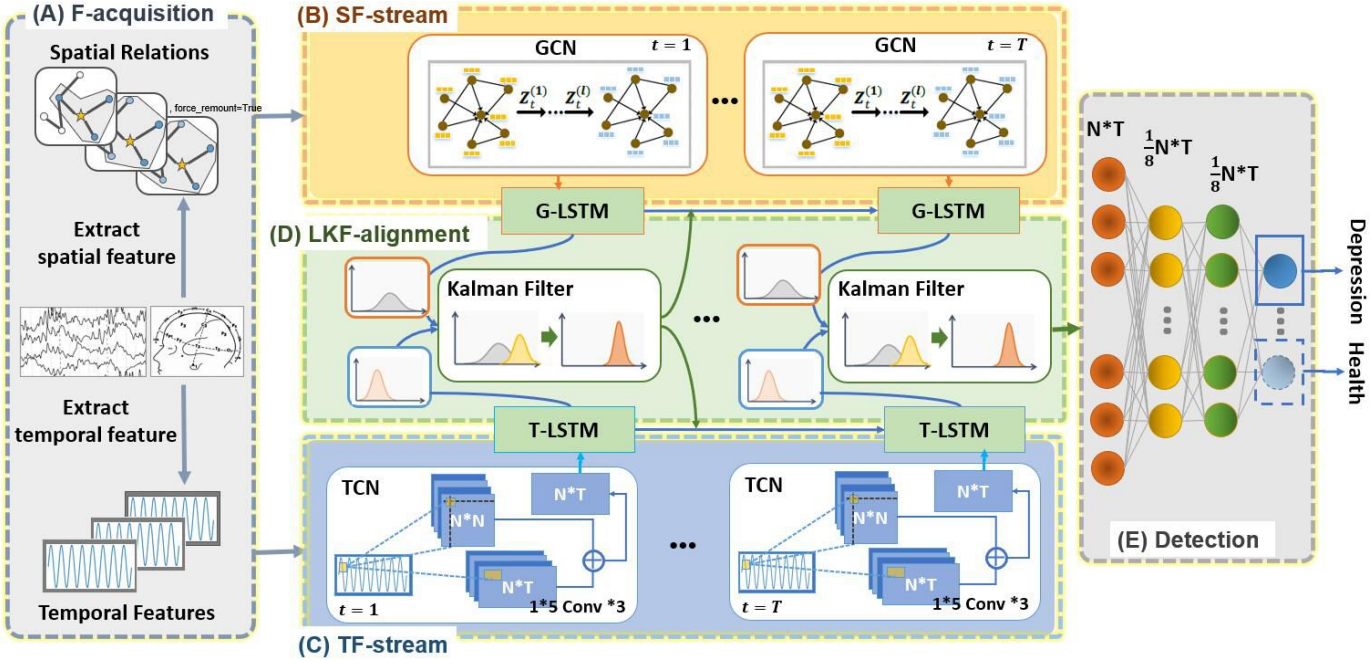

Fig. 1. LightK-DSGCN model framework. **(A)** The F-acquisition module extracts spatial and temporal features from raw EEG data, producing the k-hop adjacent matrix and temporal features **X**. **(B)** The SF-stream module integrates GCN and LSTM to encode spatial relations. It takes the k-hop adjacent matrix and temporal features **X** as input from the F-acquisition module. **(C)** The TF-stream module embeds temporal relations using integrated TCN and LSTM. It takes the temporal features **X** extracted from the F-acquisition module as input. **(D)** The LKF-alignment module denoises and merges feature embeddings from the SF-stream and TF-stream modules using a lightweight Kalman filter and LSTM layers. The output consists of integrated spatiotemporal feature vectors. **(E)** The detection module generates detection results based on the feature vectors obtained from module (D).

neighbors, the adjacency matrix is self-multiplied k times. The resulting matrices $A, A^2, ..., A^k$ from the 1 to the kth power of the adjacency matrix are stacked together, forming an $N \times (k \times N)$ matrix. Next, row-wise normalization is applied to the stacked matrix, taking into account the spatial relations of each node independently. The graph convolution feature $GC_t \in \mathbb{R}^{N \times N}$ is computed as (6).

$$GC_t = (W_{gc} \odot norm([A \oplus A^2 \oplus ... \oplus A^k])X_t \tag{6}$$

Where $W_{gc} \in \mathbb{R}^{N \times (k*N)}$ represents a trainable weight matrix for the elements in the adjacency matrices. $X_t \in \mathbb{R}^N$ refers to the input data at timestamp t. $\oplus$ denotes a vertical concatenation of matrices, and the row-wise normalization, denoted as $norm$, is applied to $A \oplus A^2 \oplus ... \oplus A^k$.

The output graph convolution features of each timestamp are then fed into an LSTM layer. For the recurrent unit of timestamp t, the input gate i_t, the output gate o_t, the forget fate f_t, and the memory cell state $\tilde{C}_t$ are computed as (7)-(10).

$$i_t = \sigma(W_i[H_{t-1} \oplus GC_t] + b_i) \tag{7}$$

$$o_t = \sigma(W_o[H_{t-1} \oplus GC_t] + b_o) \tag{8}$$

$$f_t = \sigma(W_f[H_{t-1} \oplus GC_t] + b_f) \tag{9}$$

$$\tilde{C}_t = tanh(W_c[H_{t-1} \oplus GC_t] + b_c) \tag{10}$$

Where W_i, W_o, W_f, and W_c are weights of the matrices and b_i, b_o, b_f, adn b_c are bias. H_{t-1} is the hidden state of the last timestamp. Both of H_{t-1} and GC_t are of $\mathbb{R}^{N \times N}$ features. σ is the sigmoid activation function and $tanh$ is the tangent activation function.

To address varying weights of relations among nodes at each timestamp, we introduce a cell state gate into the original LSTM model. The cell state gate is defined as shown in (11).

$$C_t^* = (W_N \odot norm([A \oplus A^2 \oplus ... \oplus A^k])C_{t-1} \tag{11}$$

Where W_N is the matrix used to re-weight the impact of each node's neighbors. The final cell state and hidden state of timestamp t is computed as (12) and -(13).

$$C_t = f_t \odot C_{t-1}^* + i_t \odot \tilde{C}_t \tag{12}$$

$$H_t = o_t \odot tanh(C_t) \tag{13}$$

C. TF-stream module

As mentioned in Section IV-A, the temporal features extracted at each point serve as input to the TF-stream module, which combines TCN and LSTM. The TF-stream module begins with a dilation convolution module that incorporates original data with multi-scale features and temporal correlations.

The dilation convolution module comprises two branches. The first branch smoothens the input temporal feature with three dilation convolution layers. Consider an EEG signal sample $\mathbf{X} \in \mathbb{R}^{F \times N \times T}$ with F original features, N channels, and T timestamps. The convolution layers within this branch retain the number of timestamps while mixing the features

ofeach timestamp with its neighboring timestamps, i.e., $\mathbf{X}_{c_i} = conv(\mathbf{X}_{c_{i-1}})$, $\mathbf{X}_{c_{i-1}} \in \mathbb{R}^{F_{c_{i-1}} \times N \times T}$, $\mathbf{X}_{c_i} \in \mathbb{R}^{F_{c_i} \times N \times T}$, where $F_{c_i} > F$ is the number of latent features in convolution layer i, and $\mathbf{X}_{c_0} = \mathbf{X}$. The feature map can then be considered as features extracted from segments of timestamps.

The second branch computes node-wise correlations for each timestamp using the expanded features. Specifically, the features are first normalized to be distributed as standard normal distributions. Therefore, the node-wise correlation is analogous to Pearson correlation over temporal segments, given the Pearson correlation equation $X_{pc}(t,(i,j)) = \frac{\mathbb{E}[(X_{t,i}-\bar{X_{t,i}})(X_{t,j}-\bar{X_{t,j}})]}{\sigma_{X_{t,i}}\sigma_{X_{t,j}}}$ and $X_{t,i} \sim \mathcal{N}(0,1)$, $X_{t,j} \sim \mathcal{N}(0,1)$. The effectiveness of Pearson correlation among channels is proved to be effective in [21]. However, we employ local, adaptive features instead of global, statistical features.

Finally, $\mathbf{X}_{c_3} \in \mathbb{R}^{F_{c_3} \times N \times T}$ and $\mathbf{X}_{pc} \in \mathbb{R}^{N \times N \times T}$ are concatenated along the feature dimension and reshaped into a $F \times N \times T$ tensor for next step.

Following the dilation TCN module, the resulting output is fed into an LSTM layer. The LSTM implementation is the same as described in Section IV-B.

D. LKF-alignment module

As mentioned, the EEG data is analyzed separately for its temporal and spatial characteristics at each time point. However, this separation of learning can result in slight disparities between the temporal and spatial relationships within the data. In this module, we employ a lightweight Kalman filter approach (described in Section III) to address this issue. The purpose is to minimize discrepancies between the statistical distributions of the feature embeddings generated by the TF-stream and SF-stream modules.

The lightweight Kalman filter approach avoids computationally intensive matrix multiplications typically found in traditional filters. Its objective is to scale the output of both the TF-stream and SF-stream modules at each timestamp with the corresponding variance of their outcomes. This scaling ensures a more consistent alignment between the temporal and spatial feature embeddings. Based on Equation 5, the merged hidden state is formulated as (14).

$$H_{kf,t} = \frac{c * Var_{tc,t} * H_{gc,t} + (1-c) * Var_{gc,t} * H_{tc,t}}{Var_{tc,t} + Var_{gc,t} + c} \tag{14}$$

where w is a trained parameter in the model. $H_{kf,t}$ is the merged hidden state. $Var_{gc,t}$ and $Var_{tc,t}$ are the variances of the hidden states of the TF-stream module and the SF-stream module, respectively. $H_{gc,t}$ and $H_{tc,t}$ are hidden states of the TF-stream module and the SF-stream module. The hidden states of these two modules are then updated as (15) and (16).

$$H'_{gc,t} = (H_{gc,t} - H_{kf,t}) \oplus H_{kf,t} \tag{15}$$

$$H'_{tc,t} = (H_{tc,t} - H_{kf,t}) \oplus H_{kf,t} \tag{16}$$

Finally, The output of the TF-stream module and the SF-stream module is merged with the historical data.

E. Detection module

After merging the spatiotemporal features, the next and final step is to choose a classifier for detecting depression in EEG signals.

We design a three-layer MLP with fully connected layers to serve as the classifier. The MLP processes the input data, specifically, the merged spatiotemporal features obtained from the LKF-alignment module. The MLP generates detection results at the output layer by learning from this input data. These results indicate the presence or absence of depression based on the learned patterns from the input data.

V. Experiment

A. Dataset

The study utilized data from two publicly available datasets: MODMA [35] and EDRA [21].

The MODMA dataset included a total of 53 participants, with 24 individuals diagnosed with Major Depressive Disorder (MDD) and 29 individuals without depression serving as normal controls. To assess depression, all subjects completed the Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) [37], with scores above 5 indicating depression. EEG signals were recorded for a duration of five minutes while the participants were in a resting state with their eyes closed. The EEG signals were captured using 128 channel electrodes placed according to the standard international 10/20 system, and the sampling frequency was set at 250 Hz.

The EDRA dataset included EEG recordings from 50 individuals, divided into two groups: 26 individuals classified as high-risk for depression and 24 individuals classified as low-risk controls. EEG signals were obtained using a 62-electrode international standard cap positioned above the right eye. The signals were sampled at a frequency of 500 Hz.

B. Evaluation metrics

We used multiple metrics to evaluate the performance of our proposed model, including Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE), as defined in (17)-(19).

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{17}$$

$$SEN = \frac{TP}{TP + FN} \tag{18}$$

$$SPE = \frac{TN}{FP + TN} \tag{19}$$

Where TP represents the count of accurately classified depression samples, FN denotes the count of depression samples that were misclassified as healthy ones, FP denotes the count of healthy instances that were misclassified as depression cases, and TN indicates the count of correctly classified healthy cases.

C. Experimental settings

The experiments were conducted in a 64-bit Linux system with an Intel(R) Core(TM) i7-6800k processor and 62GB of memory. The data was divided into training, validation, and test sets in the ratio of 8:1:1. And we conducted a ten-fold cross-validation. Table I shows the values of the parameters in our proposed framework.

TABLE I
PARAMETER SETTINGS IN OUR KFST-GNN MODEL.

Parameters	Description	Value
lr	Learning rate of our framework	2e-5
epoch	Number of epochs to train	500
hidden	Number of units in hidden layer	32, 64
dropout	Dropout rate	0.1
weight_decay	Weight on embedding matrix	1e-4
k	k-hop neighbors for GCN	4
kernel size	kernel size along the time axis for TCN	5
batch size	batch size in training process	64
optimizer	optimizer in training process	RMSprop

D. Baselines

To evaluate the effectiveness of our model, we compared LightK-DSGCN with the models proposed in the other 19 works. These models encompass a variety of categories, including three basic machine learning models, two basic deep learning models, six state-of-the-art (SOTA) machine learning models, and eight SOTA deep learning models. To maintain fairness in our performance evaluation, we extracted experiment results directly from the papers introducing these SOTA models.

Among the baselines, SVM, LR, MLP, GCN, and GAT are classical algorithms to classify depression-related features. The success of the machine learning models in EEG classification lies in their adept utilization of sophisticated pre-processing methods. These methods empower extracting intricate and abstract feature representations from EEG signals, leading to improved identification of depression-related patterns. ChSel+SVM [22] performs channel selection and SVM-based for classification. MBFCN [31] adopts multi-layer brain functional connectivity networks. FC+CFS+RF [19] employs the random forest (RF) for classification with a functional connectivity network and featured selection. PSD+MF [22] integrates multimodal features with power spectrum density, and then does classification. Node2vec+ED [16] uses the Node2vec algorithm for feature extraction, and classifies depressed patients and healthy controls with the Euclidean distance. FC+LASSO+SVM [21] builds a detection model with a functional connectivity graph and LASSO operator, and uses SVM for classification.

The deep learning methods can construct brain graphs and extract local or global features, which leads to enhanced performance compared to traditional machine learning techniques. Furthermore, a growing body of evidence highlights the extraction of temporal or spatial features from EEG signals for identifying patterns associated with depression. This extraction provides valuable insights into the underlying neural mechanisms of depression. Specifically, CNN+GA+GRU [29] combines CNN, GRU, and GA to extract features selected from brain networks. MS^2-GNN [32] is a multimodal MDD detection method incorporating modal-shared and modal-specific GNNs. ST-GCN [33] is a spatial-temporal graph convolutional network for brain functional connectivity graphs. DeepConvNet [30] is an effective and versatile framework for deep convolutional neural networks. ShallowConvNet [30] is a shallow architecture of convolutional neural networks. SparNet [23] employs a sparse convolutional neural network to learn EEG frequency features and performs depression detection. EEGNet [24] is an EEG-based brain-computer interface framework based on a compact convolutional neural network. SGP-SL [12] employs self-attention graph pooling and soft labels.

E. Results Analysis

In this section, we compared the performance of our LightK-DSGCN with baselines. As shown in Table II and Table III.

TABLE II
PERFORMANCE OF LIGHTK-DSGCN AGAINST COMPARATIVE APPROACHES ON THE MODMA DATASET. BOLDED AND BLUE RESULTS INDICATE THE BEST AND THE SECOND.

Method	Acc(%)	Sen(%)	Spe(%)
SVM	70.16	71.05	69.38
LR	82.16	81.05	82.38
MLP	91.02	91.53	89.38
GCN	95.37	95.82	94.56
GAT	94.31	93.64	94.92
ChSel+SVM [22]	81.60	-	-
MBFCN [31]	92.86	-	-
FC+CFS+RF [19]	93.20	93.90	92.80
PSD+MF [22]	92.13	93.94	89.63
Node2vec+ED [16]	88.80	-	-
CNN+GA+GRU [29]	89.63	90.24	89.63
MS^2-GNN [32]	86.49	-	-
ST-GCN [33]	93.85	-	-
DeepConvNet [30]	94.17	93.63	94.58
SparNet [23]	94.37	95.07	93.66
EEGNet [24]	92.90	92.07	93.56
SGP-SL [12]	84.91	-	-
FC+LASSO+SVM [21]	**97.43**	**96.99**	97.77
LightK-DSGCN (our)	97.22	96.57	**97.86**

TABLE III
PERFORMANCE OF LIGHTK-DSGCN AGAINST COMPARATIVE APPROACHES ON THE EDRA DATASET. BOLDED AND BLUE RESULTS INDICATE THE BEST AND THE SECOND.

Method	Acc(%)	Sen(%)	Spe(%)
SVM	68.59	67.83	69.05
LR	80.16	79.53	81.26
MLP	89.53	90.25	88.6
GCN	93.86	94.52	93.28
GAT	92.51	92.04	93.65
ShallowConvNet	84.69	80.80	82.50
EEGNet	94.67	92.33	93.70
DeepConvNet [30]	95.94	94.68	95.17
FC+LASSO+SVM [21]	97.33	98.68	96.06
LightK-DSGCN (our)	**99.41**	**99.29**	**99.60**

Overall, LightK-DSGCN outperforms other baseline models on both of the two datasets, especially on the EDRA dataset.

Specifically, our method improves accuracy by about 20% compared to the basic machine learning-based methods. This improvement can be attributed to the limitations of traditional machine learning approaches like SVM, LR, and ChSel+SVM, which rely on straightforward extraction of linear or non-linear features and struggle to effectively learn complex spatial features. Compared to deep learning baseline methods, our approach outperforms them with remarkable results. For example, our practice demonstrates an improvement in accuracy of approximately 5% compared to GCN.

Our method stands out with a 2% increase in EDRA while showing a similar performance compared to the most recent advancements in machine learning, i.e., FC+LASSO+SVM. Two factors could potentially explain this result. Firstly, MODMA has twice the number of electrodes compared to EDRA, and this increment in the number of electrodes may weaken the capabilities of the Graph Neural Network (GNN). Secondly, the input data for LightK-DSGCN is notably noisier compared to FC+LASSO+SVM, which might lead to a slight decrease in the performance of LightK-DSGCN. Nevertheless, it's worth highlighting that, on average, LightK-DSGCN exhibits the best performance. Besides, the other methods often focus on static functional connectivity while overlooking dynamic changes. In contrast, our approach harnesses the power of dynamic graph neural networks to simultaneously explore spatiotemporal features, enabling a more detailed identification of representative patterns in individuals with depression. Additionally, our LightK-DSGCN ensures the restoration of normal patterns while minimizing the influence of noises. Our approach provides a more comprehensive and accurate analysis of depression-related patterns.

F. Ablation Experiments

1) ***Setting 1. Effect of different modules****:* The purpose of these experiments is to systematically analyze the impact of removing specific modules and features on the overall performance of our model.

- **Variant a**. LightK-DSGCN without SF-stream and LKF-alignment module.
- **Variant b**. LightK-DSGCN without TF-stream and LKF-alignment module.
- **Variant c**. LightK-DSGCN without LKF-alignment module.
- **Variant d**. LightK-DSGCN without TCN module.
- **Variant e**. LightK-DSGCN (our)

As shown in Fig. 2, "Variant e" outperforms the others, indicating the positive impact of all modules on our model's detection performance. "Variant a, b, c" shows a considerable drop in performance, underscoring the pivotal role of synergizing LKF-alignment and the dual-stream modules. While the performance of "Variant d" is comparable, it notably falls behind "Variant e", suggesting that the TCN module also makes a meaningful contribution to the overall performance. The results show that the spatial information is important while the smoothened temporal features are also useful. Despite not using both spatial and temporal features simultaneously, the LKF-alignment module's denoising capability enables the model to learn effective feature embeddings, resulting in minimal performance degradation.

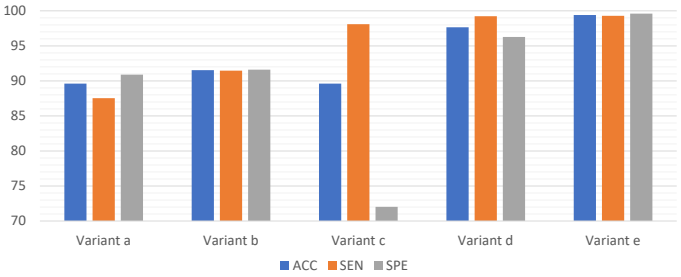

Fig. 2. Ablation study results for different modules on EDRA dataset.

2) ***Setting 2. Effect of the number of neighbors*** K*:* Here, we conducted experiments to investigate the influence of different values of K on the performance of our model.

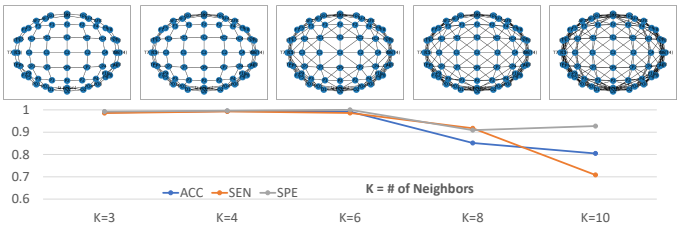

Fig. 3. The influence of different values of K on the accuracy in our LightK-DSGCN model over the EDRA dataset. The upper section showcases the channel layouts corresponding to $K = 3, 4, 6, 8, 10$, presented from left to right.

As illustrated in Fig. 3, when K is set to 4, the model achieves the highest accuracy performance. This can be attributed to the fact that a large value of K can introduce irrelevant connections between channels, while a small value of K may lead to the loss of essential correlations.

The optimization process of K consists of two stages. Initially, with a step size of 2 within the range of [2,10], we drew the graph layouts with edges. Notably, for $K = 2$, the resulting subgraph was disconnected, prompting us to replace it with $K = 3$. Subsequently, we conducted experiments with different values of K. We observed that the model's performance demonstrated sensitivity to the chosen value of K. However, again, we stress that selecting the appropriate value for K can be accomplished easily by visually inspecting the layout of the constructed graph, i.e., the optimal K is acquired when each channel is exclusively connected to adjacent channels in the up, down, left, and right directions.

VI. Conclusion

This paper introduces LightK-DSGCN, a novel depression prediction model. LightK-DSGCN combines dual-stream graph neural networks to explore spatiotemporal features and utilizes time series decomposition and a lightweight Kalman filter for denoising. The EEG signals are decomposed into

temporal and spatial features, which are then processed separately using a dilation temporal convolutional network and a functional-connectivity-based graph convolutional network, respectively. The model's effectiveness is demonstrated on two depression detection datasets, outperforming all baseline models.

REFERENCES

[1] WHO, "Depressive disorder." https://www.who.int/news-room/fact-sheets/detail/depression, 2023.

[2] J. R. W. T. D. Kaiser, R. H.and Andrews-Hanna and D. A. Pizzagalli, "Large-scale network dysfunction in major depressive disorder: A meta-analysis of resting-state functional connectivity," *Jama Psychiatry*, vol. 72, no. 6, pp. 603–611, 2015.

[3] M. M. Sanchez, L. Borden, N. Alam, A. Noroozi, M. Ravan, P. Flor-Henry, and G. Hasey, "A machine learning algorithm to discriminating between bipolar and major depressive disorders based on resting eeg data," in *Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, ser. EMBC'22, July 11-15 2022, pp. 2635–2638.

[4] J. Radua, A. Pertusa, and N. Cardoner, "Climatic relationships with specific clinical subtypes of depression," *Psychiatry Research*, vol. 175, no. 4, pp. 217–220, 2010.

[5] W. Li, H. Wang, and L. Zhuang, "Gcns-fsmi:eeg recognition of mentalillness based on fine-grained signal features and graph mutual information maximization," *Expert Systems With Applications*, vol. 228, p. 120227, 2023.

[6] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, p. 199–222, 2004.

[7] C. Li, C. Zheng, and C. Tai, "Detection of ecg characteristic points using wavelet transforms," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 2002.

[8] A. Seal, R. Bajpai, and J. Agnihotri, "Deprnet: A deep convolution neural network framework for detecting depression using eeg," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[9] M. Niu, J. Tao, and B. Liu, "Automatic depression level detection via lp-norm pooling," in *Proceedings of the 20th Annual Conference of the International Speech Communication Association*, ser. ISCA'19, September 15-19 2019, pp. 4559–4563.

[10] J. Yang, W.-S. Zheng, Q. Yang, Y.-C. Chen, and Q. Tian, "Spatial-temporal graph convolutional network for video-based person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, ser. CVPR '20, June 14-19 2020, pp. 3289–3299.

[11] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–32, 2021.

[12] T. Chen, Y. Guo, S. Hao, and R. Hong, "Exploring self-attention graph pooling with eeg-based topological structure and soft label for depression detection," *IEEE transactions on affective computing*, vol. 13, no. 4, pp. 2106–2118, 2022.

[13] J. Zhu, C. Jiang, J. Chen, X. Lin, R. Yu, X. Li, and B. Hu, "Eeg based depression recognition using improved graph convolutional neural network," *Computers in Biology and Medicine*, vol. 148, p. 105815, 2022.

[14] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, and C. Leiserson, "Evolvegcn: Evolving graph convolutional networks for dynamic graphs," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 5363–5370.

[15] M. Liu, Z. Tu, T. Su, X. Wang, X. Xu, and Z. Wang, "Behaviornet: A fine-grained behavior-aware network for dynamic link prediction," *ACM Transactions on the web*, 2023.

[16] S. Soni, A. Seal, A. Yazidi, and O. Krejcar, "Graphical representation learning-based approach for automatic classification of electroencephalogram signals in depression," *Computers in Biology and Medicine*, vol. 145, p. 105420, 2022.

[17] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal eeg data for depression recognition," *Information Fusion*, vol. 59, pp. 127–138, 2020.

[18] M. Saeedi, A. Saeedi, and A. Maghsoudi, "Major depressive disorder assessment via enhanced k-nearest neighbor method and eeg signals," *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 1007–1018, 2020.

[19] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, "A major depressive disorder classification framework based on eeg signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis," *Journal of Neuroscience Methods*, vol. 358, p. 109209, 2021.

[20] I.-M. Spyrou, C. Frantzidis, C. Bratsas, I. Antoniou, and P. D. Bamidis, "Geriatric depression symptoms coexisting with cognitive decline: a comparison of classification methodologies," *Biomedical Signal Processing and Control*, vol. 25, pp. 118–129, 2016.

[21] L. Yang, X. Wei, F. Liu, X. Zhu, and F. Zhou, "Automatic feature learning model combining functional connectivity network and graph regularization for depression detection," *Biomedical Signal Processing and Control*, vol. 82, p. 104520, 2023.

[22] J. Shen, X. Zhang, X. Huang, M. Wu, J. Gao, D. Lu, Z. Ding, and B. Hu, "An optimal channel selection for eeg-based depression detection via kernel-target alignment," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2545–2556, 2020.

[23] X. Deng, X. Fan, X. Lv, and K. Sun, "Sparnet: A convolutional neural network for eeg space-frequency feature learning and depression discrimination," *Frontiers in Neuroinformatics*, vol. 16, p. 914823, 2022.

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

[25] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated eeg-based screening of depression using deep convolutional neural network," *Computer methods and programs in biomedicine*, vol. 161, pp. 103–113, 2018.

[26] X. Li, R. La, Y. Wang, J. Niu, S. Zeng, S. Sun, and J. Zhu, "Eeg-based mild depression recognition using convolutional neural network," *Medical & biological engineering & computing*, vol. 57, pp. 1341–1352, 2019.

[27] G. Sharma, A. Parashar, and A. M. Joshi, "Dephnn: a novel hybrid neural network for electroencephalogram (eeg)-based screening of depression," *Biomedical signal processing and control*, vol. 66, p. 102393, 2021.

[28] X. Song, D. Yan, L. Zhao, and L. Yang, "Lsdd-eegnet: An efficient end-to-end framework for eeg-based depression detection," *Biomedical Signal Processing and Control*, vol. 75, p. 103612, 2022.

[29] W. Liu, K. Jia, Z. Wang, and Z. Ma, "A depression prediction algorithm based on spatiotemporal feature of eeg signal," *Brain Sciences*, vol. 12, no. 5, p. 630, 2022.

[30] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[31] X. Sun, X. Zheng, Y. Xu, L. Cui, and B. Hu, "Major depressive disorder recognition and cognitive analysis based on multi-layer brain functional connectivity networks," *arXiv preprint arXiv:2111.01351*, 2021.

[32] T. Chen, R. Hong, Y. Guo, S. Hao, and B. Hu, "Ms ˆ{2}-gnn: Exploring gnn-based multimodal fusion network for depression detection," *IEEE Transactions on Cybernetics*, 2022.

[33] H. Wu, J. Liu, and Y. Zhao, "Eeg-based depression identification using a deep learning model," in *Proceedings of the IEEE 6th Conference on Information and Communication Technology (CICT)*, ser. CICT'22. IEEE, November 18-20 2022, pp. 1–5.

[34] G. Lu, W. Ouyang, D. Xu, X. Zhang, Z. Gao, and M.-T. Sun, "Deep kalman filtering network for video compression artifact reduction," in *Proceedings of the European Conference on Computer Vision*, ser. ECCV '18, September 8 - 14 2018, pp. 568–584.

[35] H. Cai, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, Q. Zhao *et al.*, "Modma dataset: a multi-modal open dataset for mental-disorder analysis," *arXiv preprint arXiv:2002.09283*, 2020.

[36] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of 5th International Conference on Learning Representations*, ser. ICLR '17, April 24-26 2017, pp. 1–14.

[37] R. L. Spitzer, K. Kroenke, J. B. Williams, and B. Löwe, "A brief measure for assessing generalized anxiety disorder: the gad-7," *Archives of internal medicine*, vol. 166, no. 10, pp. 1092–1097, 2006.