```
import albumentations as A
import gc
import matplotlib.pyplot as plt
import math
import multiprocessing
import numpy as np
import os
import pandas as pd
import random
import time
import timm
import torch
import torch.nn as nn
from pathlib import Path


from albumentations.pytorch import ToTensorV2
from glob import glob
from torch.utils.data import DataLoader, Dataset
from tqdm import tqdm
from typing import Dict, List

from sklearn.model_selection import KFold, GroupKFold
from skimage.transform import resize
from torch.optim.lr_scheduler import OneCycleLR
import torch.nn.functional as F
import logging
import functools
import pywt

os.environ["CUDA_VISIBLE_DEVICES"] = "0,1"
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
print('Using', torch.cuda.device_count(), 'GPU(s)')



class config:
    AMP = True
    BATCH_SIZE_TRAIN = 8
```

```
    BATCH_SIZE_VALID = 8
    EPOCHS = 16
    FOLDS = 4
    FREEZE = False
    GRADIENT_ACCUMULATION_STEPS = 1
    MAX_GRAD_NORM = 1e7
    MODEL = "tf_efficientnet_b0"
    NUM_FROZEN_LAYERS = 39
    NUM_WORKERS = 0 # multiprocessing.cpu_count()
    PRINT_FREQ = 20
    SEED = 20
    TRAIN_FULL_DATA = False
    VISUALIZE = False
    WEIGHT_DECAY = 0.01


class paths:

    OUTPUT_DIR = Path("./kaggle/working/")
    TRAIN_CSV = "./balanced_train.csv"
#     TRAIN_CSV = "/kaggle/input/hms-harmful-brain-activity-classification/train.csv"
    TRAIN_EEGS = "./kaggle/input/hms-harmful-brain-activity-classification/train_eegs"
    TRAIN_SPECTROGRAMS = "./kaggle/input/hms-harmful-brain-activity-classification/train_s
    ROOT = Path.cwd()
    INPUT = ROOT / "input"
    DATA = Path("./original_data")

    PRE_LOADED_EEGS = './kaggle/working/brain-eeg/eeg_specs.npy'
    PRE_LOADED_SPECTROGRAMS = './kaggle/working/brain-spectrograms/specs.npy'
    PRE_LOADED_Wavelets = './kaggle/working/brain-wavelets/specs.npy'

log_filename = paths.ROOT/'new_version_training_record.log'

logging.basicConfig(filename=log_filename, level=logging.INFO,
                    format='%(asctime)s %(levelname)s %(message)s', datefmt='%Y-%m-%d %H:%
def log_time(func):
    """warpper for logging running time"""

    @functools.wraps(func)
```

```
    def wrapper(*args, **kwargs):
        start_time = time.time()
        result = func(*args, **kwargs)
        end_time = time.time()
        logging.info(f"{func.__name__} took {end_time - start_time:.4f} seconds.")
        print(f"{func.__name__} took {end_time - start_time:.4f} seconds.")
        return result

    return wrapper


class AverageMeter(object):
    """Computes and stores the average and current value"""
    def __init__(self):
        self.reset()

    def reset(self):
        self.val = 0
        self.avg = 0
        self.sum = 0
        self.count = 0

    def update(self, val, n=1):
        self.val = val
        self.sum += val * n
        self.count += n
        self.avg = self.sum / self.count


def asMinutes(s: float):
    "Convert to minutes."
    m = math.floor(s / 60)
    s -= m * 60
    return '%dm %ds' % (m, s)


def timeSince(since: float, percent: float):
    now = time.time()
    s = now - since
```

```
    es = s / (percent)
    rs = es - s
    return '%s (remain %s)' % (asMinutes(s), asMinutes(rs))



def plot_spectrogram(spectrogram_path: str):
    """
    Source: https://www.kaggle.com/code/mvvppp/hms-eda-and-domain-journey
    Visualize spectogram recordings from a parquet file.
    :param spectrogram_path: path to the spectogram parquet.
    """
    sample_spect = pd.read_parquet(spectrogram_path)

    split_spect = {
        "LL": sample_spect.filter(regex='^LL', axis=1),
        "RL": sample_spect.filter(regex='^RL', axis=1),
        "RP": sample_spect.filter(regex='^RP', axis=1),
        "LP": sample_spect.filter(regex='^LP', axis=1),
    }

    fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(15, 12))
    axes = axes.flatten()
    label_interval = 5
    for i, split_name in enumerate(split_spect.keys()):
        ax = axes[i]
        img = ax.imshow(np.log(split_spect[split_name]).T, cmap='viridis', aspect='auto',
        cbar = fig.colorbar(img, ax=ax)
        cbar.set_label('Log(Value)')
        ax.set_title(split_name)
        ax.set_ylabel("Frequency (Hz)")
        ax.set_xlabel("Time")

        ax.set_yticks(np.arange(len(split_spect[split_name].columns)))
        ax.set_yticklabels([column_name[3:] for column_name in split_spect[split_name].co
        frequencies = [column_name[3:] for column_name in split_spect[split_name].columns
        ax.set_yticks(np.arange(0, len(split_spect[split_name].columns), label_interval))
        ax.set_yticklabels(frequencies[::label_interval])
    plt.tight_layout()
```

```
    plt.show()

@log_time
def seed_everything(seed: int):
    random.seed(seed)
    np.random.seed(seed)
    torch.manual_seed(seed)
    os.environ['PYTHONHASHSEED'] = str(seed)


def sep():
    print("-"*100)


class CustomDataset(Dataset):
    def __init__(
        self, df: pd.DataFrame, config,
        augment: bool = False, mode: str = 'train',
        specs: Dict[int, np.ndarray] = None,
        eeg_specs: Dict[int, np.ndarray] = None,
        wavelets_spectrograms: Dict[int, np.ndarray] = None
    ):
        self.df = df
        self.config = config
        self.batch_size = self.config.BATCH_SIZE_TRAIN
        self.augment = augment
        self.mode = mode
        self.spectrograms = specs if specs is not None else {}
        self.eeg_spectrograms = eeg_specs if eeg_specs is not None else {}
        self.wavelets_spectrograms = wavelets_spectrograms if wavelets_spectrograms is not

    def __len__(self):
        """
        Denotes the number of batches per epoch.
        """
        return len(self.df)

    def __getitem__(self, index):
        """
```

```
        Generate one batch of data.
        """
        X, y = self.__data_generation(index)
        if self.augment:
            X = self.__transform(X)
        return torch.tensor(X, dtype=torch.float32), torch.tensor(y, dtype=torch.float32)

    def log_and_Standarize(self,img):
        # Log transform spectogram
            img = np.clip(img, np.exp(-4), np.exp(8))
            img = np.log(img)

            # Standarize per image
            ep = 1e-6
            mu = np.nanmean(img.flatten())
            std = np.nanstd(img.flatten())
            img = (img - mu) / (std + ep)
            img = np.nan_to_num(img, nan=0.0)
            return img

    def __data_generation(self, index):
        """
        Generates data containing batch_size samples.
        """
        X = np.zeros((128, 256, 12), dtype='float32')
        y = np.zeros(6, dtype='float32')
        img = np.ones((128,256), dtype='float32')
        row = self.df.iloc[index]
        if self.mode=='test':
            r = 0
        else:
            r = int((row['min'] + row['max']) // 4)

        for region in range(4):
            img = self.spectrograms[row.spectrogram_id][r:r+300, region*100:(region+1)*100
            img = self.log_and_Standarize(img)
            X[14:-14, :, region] = img[:, 22:-22] / 2.0

        img = self.eeg_spectrograms[row.eeg_id]
```

```
        img = img.to_numpy()
        img = self.log_and_Standarize(img)
        img = resize(img, (128, 256, 4))
        X[:, :, 4:8] = img

        # Combine wavelet features
        img = self.wavelets_spectrograms[row.spectrogram_id]
        img = self.log_and_Standarize(img)
        img = resize(img, (128, 256,4))
        X[:, :, 8:12] = img


        if self.mode != 'test':
            y = row[label_cols].values.astype(np.float32)

        return X, y

    def __transform(self, img):
        transforms = A.Compose([
            A.HorizontalFlip(p=0.5),
        ])
        return transforms(image=img)['image']


class CustomModel(nn.Module):
    def __init__(self, config, num_classes: int = 6, pretrained: bool = True):
        super(CustomModel, self).__init__()
        self.USE_KAGGLE_SPECTROGRAMS = True
        self.USE_EEG_SPECTROGRAMS = False
        self.USE_WAVELET_SPECTROGRAMS = False
        self.model = timm.create_model(
            config.MODEL,
            pretrained=pretrained,
            drop_rate = 0.1,
            drop_path_rate = 0.2,
        )
        # add code on logging parameter
        logging.info("config.MODEL: {}".format(config.MODEL))
        logging.info("USE_KAGGLE_SPECTROGRAMS: {}".format(self.USE_KAGGLE_SPECTROGRAMS))
```

```
        logging.info("USE_EEG_SPECTROGRAMS: {}".format(self.USE_EEG_SPECTROGRAMS))
        logging.info("USE_WAVELET_SPECTROGRAMS: {}".format(self.USE_WAVELET_SPECTROGRAMS))

        if config.FREEZE:
            for i,(name, param) in enumerate(list(self.model.named_parameters())\
                                             [0:config.NUM_FROZEN_LAYERS]):
                param.requires_grad = False

        self.features = nn.Sequential(*list(self.model.children())[:-2])
        self.custom_layers = nn.Sequential(
            nn.AdaptiveAvgPool2d(1),
            nn.Flatten(),
            nn.Linear(self.model.num_features, num_classes)
        )

    def __reshape_input(self, x):
        """
        Reshapes input torch.Size([8, 128, 256, 12]) -> [8, 3, 512, 768] monotone image.
        """
        components = []
        if self.USE_KAGGLE_SPECTROGRAMS:
            spectograms = [x[:, :, :, i:i+1] for i in range(4)]
            components.append(torch.cat(spectograms, dim=1))
        if self.USE_EEG_SPECTROGRAMS:
            eegs = [x[:, :, :, i:i+1] for i in range(4,8)]
            eegs = torch.cat(eegs, dim=1)
            components.append(eegs)

        if self.USE_WAVELET_SPECTROGRAMS:
            wavelets = [x[:, :, :, i:i+1] for i in range(8,12)]
            wavelets = torch.cat(wavelets, dim=1)
            components.append(wavelets)

        if components:
            x = torch.cat(components, dim=2)

        x = torch.cat([x, x, x], dim=3)
        x = x.permute(0, 3, 1, 2)
        return x
```

```
    def forward(self, x):
        x = self.__reshape_input(x)
        x = self.features(x)
        x = self.custom_layers(x)
        return x


@log_time
def train_epoch(train_loader, model, criterion, optimizer, epoch, scheduler, device):
    """One epoch training pass."""
    model.train()
    criterion = nn.KLDivLoss(reduction="batchmean")
    scaler = torch.cuda.amp.GradScaler(enabled=config.AMP)
    losses = AverageMeter()
    start = end = time.time()
    global_step = 0

    # ========== ITERATE OVER TRAIN BATCHES ============
    with tqdm(train_loader, unit="train_batch", desc='Train') as tqdm_train_loader:
        for step, (X, y) in enumerate(tqdm_train_loader):
            X = X.to(device)
            y = y.to(device)
            batch_size = y.size(0)
            with torch.cuda.amp.autocast(enabled=config.AMP):
                y_preds = model(X)
                loss = criterion(F.log_softmax(y_preds, dim=1), y)
            if config.GRADIENT_ACCUMULATION_STEPS > 1:
                loss = loss / config.GRADIENT_ACCUMULATION_STEPS
            losses.update(loss.item(), batch_size)
            scaler.scale(loss).backward()
            grad_norm = torch.nn.utils.clip_grad_norm_(model.parameters(), config.MAX_GRAD

            if (step + 1) % config.GRADIENT_ACCUMULATION_STEPS == 0:
                scaler.step(optimizer)
                scaler.update()
                optimizer.zero_grad()
                global_step += 1
                scheduler.step()
```

```
            end = time.time()

            # ========== LOG INFO ==========
            if step % config.PRINT_FREQ == 0 or step == (len(train_loader)-1):
                print('Epoch: [{0}][{1}/{2}] '
                      'Elapsed {remain:s} '
                      'Loss: {loss.avg:.4f} '
                      'Grad: {grad_norm:.4f}  '
                      'LR: {lr:.8f}  '
                      .format(epoch+1, step, len(train_loader),
                              remain=timeSince(start, float(step+1)/len(train_loader)),
                              loss=losses,
                              grad_norm=grad_norm,
                              lr=scheduler.get_last_lr()[0]))

    return losses.avg

@log_time
def valid_epoch(valid_loader, model, criterion, device):
    model.eval()
    softmax = nn.Softmax(dim=1)
    losses = AverageMeter()
    prediction_dict = {}
    preds = []
    start = end = time.time()
    with tqdm(valid_loader, unit="valid_batch", desc='Validation') as tqdm_valid_loader:
        for step, (X, y) in enumerate(tqdm_valid_loader):
            X = X.to(device)
            y = y.to(device)
            batch_size = y.size(0)
            with torch.no_grad():
                y_preds = model(X)
                loss = criterion(F.log_softmax(y_preds, dim=1), y)
            if config.GRADIENT_ACCUMULATION_STEPS > 1:
                loss = loss / config.GRADIENT_ACCUMULATION_STEPS
            losses.update(loss.item(), batch_size)
            y_preds = softmax(y_preds)
            preds.append(y_preds.to('cpu').numpy())
            end = time.time()
```

```
            # ========== LOG INFO ==========
            if step % config.PRINT_FREQ == 0 or step == (len(valid_loader)-1):
                print('EVAL: [{0}/{1}] '
                      'Elapsed {remain:s} '
                      'Loss: {loss.avg:.4f} '
                      .format(step, len(valid_loader),
                              remain=timeSince(start, float(step+1)/len(valid_loader)),
                              loss=losses))

    prediction_dict["predictions"] = np.concatenate(preds)
    return losses.avg, prediction_dict

@log_time
def train_loop(df, fold,stage = 2):

    # ======== SPLIT ==========
    train_folds = df[df['fold'] != fold].reset_index(drop=True)
    valid_folds = df[df['fold'] == fold].reset_index(drop=True)
    # ---------- votes sum-------------------
    if stage == 1:
        # all data
        print("Training Stage 1: Using all data")
    elif stage == 2:
        # KL Loss < 9 data
        print("Training Stage 2: Filtering data based on KL Loss < 7.5")
        train_folds = train_folds[train_folds['kl_loss'] < 7.5]

    # ======== DATASETS ==========
    train_dataset = CustomDataset(train_folds, config, mode="train", augment=True, specs=a
    valid_dataset = CustomDataset(valid_folds, config, mode="train", augment=False, specs=


    # ======== DATALOADERS ==========
    train_loader = DataLoader(train_dataset,
                              batch_size=config.BATCH_SIZE_TRAIN,
                              shuffle=True,
                              num_workers=config.NUM_WORKERS, pin_memory=True, drop_last=T
    valid_loader = DataLoader(valid_dataset,
```

```
                            batch_size=config.BATCH_SIZE_VALID,
                            shuffle=False,
                            num_workers=config.NUM_WORKERS, pin_memory=True, drop_last=F

    # ======== MODEL ==========
    model = CustomModel(config)
    model.to(device)
    if stage == 2:
        model_path = paths.OUTPUT_DIR / f"{config.MODEL.replace('/', '_')}_fold_{fold}_sta
        model.load_state_dict(torch.load(model_path)["model"])
        model.to(device)

    optimizer = torch.optim.AdamW(model.parameters(), lr=0.1, weight_decay=config.WEIGHT_D
    scheduler = OneCycleLR(
        optimizer,
        max_lr=1e-4,
        epochs=config.EPOCHS,
        steps_per_epoch=len(train_loader),
        pct_start=0.1,
        anneal_strategy="cos",
        final_div_factor=100,
    )

    # ======= LOSS ==========
    criterion = nn.KLDivLoss(reduction="batchmean")

    best_loss = np.inf
    early_stop_threshold = 4
    improvement_count= 0
    # ====== ITERATE EPOCHS ========
    for epoch in range(config.EPOCHS):
        start_time = time.time()

        # ======= TRAIN ==========
        avg_train_loss = train_epoch(train_loader, model, criterion, optimizer, epoch, sch

        # ======= EVALUATION ==========
        avg_val_loss, prediction_dict = valid_epoch(valid_loader, model, criterion, device
        predictions = prediction_dict["predictions"]
```

```
        # ======= SCORING ==========
        elapsed = time.time() - start_time

        print(f'Epoch {epoch+1} - avg_train_loss: {avg_train_loss:.4f}  avg_val_loss: {avg

        if avg_val_loss < best_loss:
            best_loss = avg_val_loss
            logging.info(f'Epoch {epoch+1} - Save Best Loss: {best_loss:.4f} Model')
            model_save_path = paths.OUTPUT_DIR / f"{config.MODEL.replace('/', '_')}_fold_
            torch.save({'model': model.state_dict(), 'predictions': predictions}, model_sa
        else:
            improvement_count += 1
            if improvement_count >= early_stop_threshold:
                print(f"Early stopping triggered at {epoch} epochs without improvement.")
                break  # early stop

    ## TypeError: unsupported operand type(s) for +: 'WindowsPath' and 'str'
    # predictions = torch.load(paths.OUTPUT_DIR + f"/{config.MODEL.replace('/', '_')}_fold
    #                          map_location=torch.device('cpu'))['predictions']
    paths.OUTPUT_DIR.mkdir(parents=True, exist_ok=True)
    predictions = torch.load(model_save_path,
                         map_location=torch.device('cpu'))['predictions']

    valid_folds[target_preds] = predictions

    torch.cuda.empty_cache()
    gc.collect()

    return valid_folds

@log_time
def train_loop_full_data(df):
    train_dataset = CustomDataset(df, config, mode="train", augment=True,specs=all_spectro
    train_loader = DataLoader(train_dataset,
                              batch_size=config.BATCH_SIZE_TRAIN,
                              shuffle=False,
                              num_workers=config.NUM_WORKERS, pin_memory=True, drop_last=T
    model = CustomModel(config)
```

```
    model.to(device)
    optimizer = torch.optim.AdamW(model.parameters(), lr=0.1, weight_decay=config.WEIGHT_D
    scheduler = OneCycleLR(
        optimizer,
        max_lr=1e-3,
        epochs=config.EPOCHS,
        steps_per_epoch=len(train_loader),
        pct_start=0.1,
        anneal_strategy="cos",
        final_div_factor=100,
    )
    criterion = nn.KLDivLoss(reduction="batchmean")
    best_loss = np.inf
    for epoch in range(config.EPOCHS):
        start_time = time.time()
        avg_train_loss = train_epoch(train_loader, model, criterion, optimizer, epoch, sch
        elapsed = time.time() - start_time
        logging.info(f'Epoch {epoch+1} - avg_train_loss: {avg_train_loss:.4f}  time: {elap
        torch.save(
            {'model': model.state_dict()},
            paths.OUTPUT_DIR + f"/{config.MODEL.replace('/', '_')}_epoch_{epoch}.pth")
    torch.cuda.empty_cache()
    gc.collect()
    return

@log_time
def get_result(oof_df):
    kl_loss = nn.KLDivLoss(reduction="batchmean")
    labels = torch.tensor(oof_df[label_cols].values)
    preds = torch.tensor(oof_df[target_preds].values)
    preds = F.log_softmax(preds, dim=1)
    result = kl_loss(preds, labels)
    return result

@log_time
def preparing_data(df):
    train_df = df.groupby('eeg_id')[['spectrogram_id','spectrogram_label_offset_seconds']]
        'spectrogram_id':'first',
        'spectrogram_label_offset_seconds':'min'
```

```
    })
    train_df.columns = ['spectrogram_id','min']

    aux = df.groupby('eeg_id')[['spectrogram_id','spectrogram_label_offset_seconds']].agg
        'spectrogram_label_offset_seconds':'max'
    })
    train_df['max'] = aux

    aux = df.groupby('eeg_id')[['patient_id']].agg('first')
    train_df['patient_id'] = aux

    aux = df.groupby('eeg_id')[label_cols].agg('sum')
    for label in label_cols:
        train_df[label] = aux[label].values

#     train_df['total_votes'] = train_df[label_cols].sum(axis=1)
    kl = compute_kl_divergence(train_df,label_cols)
    train_df['kl_loss'] = kl

    y_data = train_df[label_cols].values
    y_data = y_data / y_data.sum(axis=1,keepdims=True)
    train_df[label_cols] = y_data

    aux = df.groupby('eeg_id')[['expert_consensus']].agg('first')
    train_df['target'] = aux

    train_df = train_df.reset_index()
    return train_df

def compute_wavelet_features(signal, wavelet='db4', level=5):
    coeffs = pywt.wavedec(signal, wavelet, level=level)
        # Extract features from wavelet coefficients instead of using wavelet coefficients
    features = []
    for coeff in coeffs:
        features.extend([np.mean(coeff), np.std(coeff)])
    return np.array(features)


@log_time
```

```python
def loading_parquet(train_df, config = config, READ_SPEC_FILES = True,READ_EEG_SPEC_FILES
    paths_spectrograms = glob(paths.TRAIN_SPECTROGRAMS + "*.parquet")
    # paths_spectrograms = glob(str(paths.TRAIN_SPECTROGRAMS / "*.parquet"))
    print(f'There are {len(paths_spectrograms)} spectrogram parquets in total path')

    if READ_SPEC_FILES:
        all_spectrograms = {}
        all_wavelet_spectrograms = {}
        spectogram_ids = train_df['spectrogram_id'].unique()
        print(f'There are {len(spectogram_ids)} spectrogram parquets in this training proc
        for spec_id in tqdm(spectogram_ids):
        # for file_path in tqdm(paths_spectograms):
            file_path = f"{paths.TRAIN_SPECTROGRAMS}/{spec_id}.parquet"
            aux = pd.read_parquet(file_path)
            spec_arr = aux.fillna(0).values[:, 1:].T.astype("float32")  # (Hz, Time) = (40
            wavelet_features = np.array([compute_wavelet_features(row, wavelet=wavelet) fo
            name = int(file_path.split("/")[-1].split('.')[0])
            # all_spectrograms[name] = aux.iloc[:,1:].values
            all_spectrograms[name] = aux.fillna(0).iloc[:,1:].values.astype("float32")
            all_wavelet_spectrograms[name] = wavelet_features
            del aux
            del wavelet_features
        os.makedirs(os.path.dirname(paths.PRE_LOADED_SPECTROGRAMS), exist_ok=True)
        os.makedirs(os.path.dirname(paths.PRE_LOADED_Wavelets), exist_ok=True)
        np.save(paths.PRE_LOADED_SPECTROGRAMS, all_spectrograms, allow_pickle=True)
        np.save(paths.PRE_LOADED_Wavelets, all_wavelet_spectrograms, allow_pickle=True)
    else:
        all_spectrograms = np.load(paths.PRE_LOADED_SPECTROGRAMS, allow_pickle=True).item
        all_wavelet_spectrograms = np.load(paths.PRE_LOADED_Wavelets, allow_pickle=True).i

    if config.VISUALIZE:
        idx = np.random.randint(0,len(paths_spectrograms))
        spectrogram_path = paths_spectrograms[idx]
        plot_spectrogram(spectrogram_path)

    # Read EEG Spectrograms
    paths_eegs = glob(paths.TRAIN_EEGS + "*.parquet")
#     paths_eegs = glob(str(paths.TRAIN_EEGS / "*.parquet"))
    print(f'There are {len(paths_eegs)} EEG spectrograms in total path')
```

```
    if READ_EEG_SPEC_FILES:
        all_eegs = {}
        eeg_ids = train_df['eeg_id'].unique()
        print(f'There are {len(eeg_ids)} EEG spectrograms in this training path')
        for eeg_id in tqdm(eeg_ids):
            file_path = f"{paths.TRAIN_EEGS}/{eeg_id}.parquet"
            eeg_spectogram =  pd.read_parquet(file_path)
            all_eegs[eeg_id] = eeg_spectogram
            del eeg_spectogram
        os.makedirs(os.path.dirname(paths.PRE_LOADED_EEGS), exist_ok=True)
        np.save(paths.PRE_LOADED_EEGS, all_eegs, allow_pickle=True)
    else:
        all_eegs = np.load(paths.PRE_LOADED_EEGS, allow_pickle=True).item()




    return all_spectrograms,all_eegs,all_wavelet_spectrograms


def plot_total_votes_vs_kl_divergence(dataframe):
    plt.figure(figsize=(10, 6))
    plt.scatter(dataframe['total_votes'], dataframe['kl_divergence'], alpha=0.6, edgecolo
    plt.title('Scatter Plot of Total Votes vs. KL Divergence')
    plt.xlabel('Total Votes')
    plt.ylabel('KL Divergence')
    plt.grid(True)
    plt.savefig("filenamez", format='png', dpi=300)
#     plt.show()


def compute_kl_divergence(data, label_cols):
    labels = data[label_cols].values + 1e-5
    labels /= labels.sum(axis=1, keepdims=True)
    kl_div = torch.nn.functional.kl_div(
        torch.log(torch.tensor(labels, dtype=torch.float)),
        torch.tensor([[1/6] * 6], dtype=torch.float),
        reduction='none'
    ).sum(dim=1).numpy()
```

```
    return kl_div

if __name__ == "__main__":
    overall_start_time = time.time()
    print(f"Log file path: {log_filename.absolute()}")
    logging.info('--------------------------------------------------')
    logging.info(f'training on local balanced data')
    logging.info(f'Into loading stage')

    target_preds = [x + "_pred" for x in ['seizure_vote', 'lpd_vote', 'gpd_vote', 'lrda_v
    label_to_num = {'Seizure': 0, 'LPD': 1, 'GPD': 2, 'LRDA': 3, 'GRDA': 4, 'Other':5}
    num_to_label = {v: k for k, v in label_to_num.items()}
    seed_everything(config.SEED)

    df = pd.read_csv(paths.TRAIN_CSV)
    label_cols = df.columns[-6:]
    print(f"Train cataframe shape is: {df.shape}")
    print(f"Labels: {list(label_cols)}")
    print(df.head())

#     plot_total_votes_vs_kl_divergence(df)
#     print(sss)
    #处理train_df，eeg_id,只保留第一个spectrogram_id，min及max spec offset，第一个patient_id等
    train_df = preparing_data(df)
    print('Train non-overlapp eeg_id shape:', train_df.shape )
    print(train_df.head())
#     train_df.to_csv('./local_train_df.csv', index=False)

    logging.info(f'Into loading stage: combine wavelet feature into X')
    logging.info(f'Into loading stage: loading single npy from local file')
    all_spectrograms,all_eegs,all_wavelet_spectrograms = loading_parquet(train_df, config



    # Validation
    gkf = GroupKFold(n_splits=config.FOLDS)
    for fold, (train_index, valid_index) in enumerate(gkf.split(train_df, train_df.target
        train_df.loc[valid_index, "fold"] = int(fold)

    print(train_df.groupby('fold').size()), sep()
```

```
print(train_df.head())

train_dataset = CustomDataset(train_df, config, mode="train",
                              specs=all_spectrograms, eeg_specs=all_eegs,wavelets_spec
train_loader = DataLoader(
    train_dataset,
    batch_size=config.BATCH_SIZE_TRAIN,
    shuffle=False,
    num_workers=config.NUM_WORKERS, pin_memory=True, drop_last=True
)
X, y = train_dataset[0]
print(f"X shape: {X.shape}")
print(f"y shape: {y.shape}")

if config.VISUALIZE:
    ROWS = 2
    COLS = 3
    for (X, y) in train_loader:
        plt.figure(figsize=(20,8))
        for row in range(ROWS):
            for col in range(COLS):
                plt.subplot(ROWS, COLS, row*COLS + col+1)
                t = y[row*COLS + col]
                img = X[row*COLS + col, :, :, 0]
                mn = img.flatten().min()
                mx = img.flatten().max()
                img = (img-mn)/(mx-mn)
                plt.imshow(img)
                tars = f'[{t[0]:0.2f}'
                for s in t[1:]:
                    tars += f', {s:0.2f}'
                eeg = train_df.eeg_id.values[row*config.BATCH_SIZE_TRAIN + row*COLS +
                plt.title(f'EEG = {eeg}\nTarget = {tars}',size=12)
                plt.yticks([])
                plt.ylabel('Frequencies (Hz)',size=14)
                plt.xlabel('Time (sec)',size=16)
        plt.show()
        break
```

```
    # #dynamic learning rate
#     EPOCHS = config.EPOCHS
#     BATCHES = len(train_loader)
#     steps = []
#     lrs = []
#     optim_lrs = []
#     model = CustomModel(config)
#     optimizer = torch.optim.AdamW(model.parameters(), lr=1e-4)
#     scheduler = OneCycleLR(
#         optimizer,
#         max_lr=1e-3,
#         epochs=config.EPOCHS,
#         steps_per_epoch=len(train_loader),
#         pct_start=0.05,
#         anneal_strategy="cos",
#         final_div_factor=100,
#     )
#     for epoch in range(EPOCHS):
#         for batch in range(BATCHES):
#             scheduler.step()
#             lrs.append(scheduler.get_last_lr()[0])
#             steps.append(epoch * BATCHES + batch)

#     max_lr = max(lrs)
#     min_lr = min(lrs)
#     print(f"Maximum LR: {max_lr} | Minimum LR: {min_lr}")
    # plt.figure()
    # plt.plot(steps, lrs, label='OneCycle')
    # plt.ticklabel_format(axis='y', style='sci', scilimits=(0,0))
    # plt.xlabel("Step")
    # plt.ylabel("Learning Rate")
    # plt.show()

    if not config.TRAIN_FULL_DATA:
        oof_df = pd.DataFrame()
        for fold in range(config.FOLDS):
            for stage in [1, 2]:
                print(f"Starting Stage {stage} Training for Fold {fold}")
                _oof_df = train_loop(train_df, fold, stage=stage)
```

```
                oof_df = pd.concat([oof_df, _oof_df])
                logging.info(f"========== Fold {fold} Stage {stage} result: {get_result(_
                print(f"========== Fold {fold} Stage {stage} result: {get_result(_oof_df)
        oof_df = oof_df.reset_index(drop=True)
        print(f"========== CV: {get_result(oof_df)} ==========")
        logging.info(f"----------------------------------------------------------------------
        oof_df.to_csv(os.path.join(paths.OUTPUT_DIR, 'oof_df.csv'), index=False)
else:
    train_loop_full_data(train_df)
```

```
Using 0 GPU(s)
Log file path: /Users/Evelyn/UOS2/tp/new_version_training_record.log
seed_everything took 0.0082 seconds.
Train cataframe shape is: (300, 15)
Labels: ['seizure_vote', 'lpd_vote', 'gpd_vote', 'lrda_vote', 'grda_vote', '
       eeg_id  eeg_sub_id  eeg_label_offset_seconds  spectrogram_id  \
0  1940666997           9                      60.0       596909244
1  2620674843           1                       4.0       800599706
2  2166673542           1                       2.0       928825124
3   839616512          19                      60.0        81278784
4   761869179           7                      60.0      1247953913

   spectrogram_sub_id  spectrogram_label_offset_seconds    label_id  \
0                   9                              60.0   917008523
1                   1                               4.0   750774119
2                   1                               2.0   991434112
3                  19                              60.0  2219696038
4                   7                              60.0  2165106858

   patient_id expert_consensus  seizure_vote  lpd_vote  gpd_vote  lrda_vote
0       57251          Seizure             3         0         0          0
1       11439          Seizure             3         0         0          0
2       21771          Seizure             3         0         0          0
3       16805          Seizure             3         0         0          0
4        4898          Seizure             3         0         0          0

   grda_vote  other_vote
0          0           0
1          0           0
2          0           0
3          0           0
4          0           0
preparing_data took 0.0175 seconds.
Train non-overlapp eeg_id shape: (287, 13)
     eeg_id  spectrogram_id     min     max  patient_id  seizure_vote  \
0   4431217      1459125071    80.0    80.0       49713           0.0
1  21054661      1067342787   140.0   428.0       37979           0.0
2  54759002      1506575594    62.0    62.0       63918           1.0
3  75373657        38412976    48.0    48.0        1851           0.0
```

```
4  86189315       525426737  3076.0  3076.0       23337           0.0

   lpd_vote  gpd_vote  lrda_vote  grda_vote  other_vote   kl_loss   target
0       1.0       0.0   0.000000        0.0    0.000000  9.939808      LPD
1       0.0       0.0   0.800000        0.0    0.200000  8.456420     LRDA
2       0.0       0.0   0.000000        0.0    0.000000  8.717875  Seizure
3       0.0       0.0   0.000000        1.0    0.000000  8.717875     GRDA
4       0.0       0.0   0.615385        0.0    0.384615  7.833665     LRDA
There are 0 spectrogram parquets in total path
There are 280 spectrogram parquets in this training process


100%|██████████████████████████████████████████| 280/280 [00:14<00:00, 19.96i


There are 0 EEG spectrograms in total path
There are 287 EEG spectrograms in this training path


100%|██████████████████████████████████████████| 287/287 [00:01<00:00, 270.66i


loading_parquet took 16.5921 seconds.
fold
0.0    72
1.0    72
2.0    72
3.0    71
dtype: int64
--------------------------------------------------------------------------------
     eeg_id  spectrogram_id     min     max  patient_id  seizure_vote  \
0   4431217      1459125071    80.0    80.0       49713           0.0
1  21054661      1067342787   140.0   428.0       37979           0.0
2  54759002      1506575594    62.0    62.0       63918           1.0
3  75373657        38412976    48.0    48.0        1851           0.0
4  86189315       525426737  3076.0  3076.0       23337           0.0

   lpd_vote  gpd_vote  lrda_vote  grda_vote  other_vote   kl_loss   target
0       1.0       0.0   0.000000        0.0    0.000000  9.939808      LPD
```

```
1       0.0       0.0   0.800000        0.0    0.200000  8.456420     LRDA
2       0.0       0.0   0.000000        0.0    0.000000  8.717875  Seizure
3       0.0       0.0   0.000000        1.0    0.000000  8.717875     GRDA
4       0.0       0.0   0.615385        0.0    0.384615  7.833665     LRDA

   fold
0   0.0
1   0.0
2   3.0
3   2.0
4   1.0
X shape: torch.Size([128, 256, 12])
y shape: torch.Size([6])
Starting Stage 1 Training for Fold 0
Training Stage 1: Using all data


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
  warnings.warn(
Train:   0%|                                          | 0/26 [00:00<?, ?train_batc
  warnings.warn(
Train:   4%|█                          | 1/26 [00:09<04:04,  9.79s/train_ba

Epoch: [1][0/26] Elapsed 0m 9s (remain 4m 4s) Loss: 1.5628 Grad: 2.7134  LR:


Train:  81%|████████████████████     | 21/26 [10:02<02:23, 28.63s/train_ba

Epoch: [1][20/26] Elapsed 10m 2s (remain 2m 23s) Loss: 1.3906 Grad: 2.4186


Train: 100%|████████████████████████| 26/26 [12:39<00:00, 29.23s/train_ba


Epoch: [1][25/26] Elapsed 12m 39s (remain 0m 0s) Loss: 1.3773 Grad: 1.8469
train_epoch took 759.9127 seconds.


Validation:  11%|██                         | 1/9 [00:09<01:16,  9.60s/valid_ba
```

```
EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.5442


Validation: 100%|████████████████████████| 9/9 [01:37<00:00, 10.79s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.4090
valid_epoch took 97.1178 seconds.
Epoch 1 - avg_train_loss: 1.3773  avg_val_loss: 1.4090  time: 857s


Train:   4%|▉                         | 1/26 [00:15<06:24, 15.36s/train_ba

Epoch: [2][0/26] Elapsed 0m 15s (remain 6m 24s) Loss: 1.6451 Grad: 3.0663  L


Train:  81%|████████████████████     | 21/26 [11:18<02:02, 24.47s/train_ba

Epoch: [2][20/26] Elapsed 11m 18s (remain 2m 41s) Loss: 1.3425 Grad: 2.3154


Train: 100%|████████████████████████| 26/26 [12:38<00:00, 29.18s/train_ba


Epoch: [2][25/26] Elapsed 12m 38s (remain 0m 0s) Loss: 1.3412 Grad: 2.5044
train_epoch took 758.7258 seconds.


Validation:  11%|██                       | 1/9 [00:09<01:15,  9.46s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 15s) Loss: 1.5443


Validation: 100%|████████████████████████| 9/9 [01:36<00:00, 10.70s/valid_ba


EVAL: [8/9] Elapsed 1m 36s (remain 0m 0s) Loss: 1.3938
valid_epoch took 96.3374 seconds.
```

```
Epoch 2 - avg_train_loss: 1.3412  avg_val_loss: 1.3938  time: 855s


Train:   4%|▉                           | 1/26 [00:08<03:29,  8.37s/train_ba

Epoch: [3][0/26] Elapsed 0m 8s (remain 3m 29s) Loss: 1.0569 Grad: 2.3634  LR


Train:  81%|████████████████████      | 21/26 [10:18<01:07, 13.41s/train_ba

Epoch: [3][20/26] Elapsed 10m 18s (remain 2m 27s) Loss: 1.2442 Grad: 2.7915


Train: 100%|████████████████████████| 26/26 [13:01<00:00, 30.05s/train_ba


Epoch: [3][25/26] Elapsed 13m 1s (remain 0m 0s) Loss: 1.2494 Grad: 2.7122  L
train_epoch took 781.3759 seconds.


Validation:  11%|██                  | 1/9 [10:28<1:23:50, 628.78s/valid_ba

EVAL: [0/9] Elapsed 10m 28s (remain 83m 50s) Loss: 1.4957


Validation: 100%|████████████████████| 9/9 [11:55<00:00, 79.48s/valid_ba


EVAL: [8/9] Elapsed 11m 55s (remain 0m 0s) Loss: 1.3152
valid_epoch took 715.3522 seconds.
Epoch 3 - avg_train_loss: 1.2494  avg_val_loss: 1.3152  time: 1497s


Train:   4%|▉                           | 1/26 [00:13<05:38, 13.52s/train_ba

Epoch: [4][0/26] Elapsed 0m 13s (remain 5m 38s) Loss: 1.1981 Grad: 2.4502  L


Train:  81%|████████████████████      | 21/26 [10:11<01:46, 21.39s/train_ba
```

```
Epoch: [4][20/26] Elapsed 10m 11s (remain 2m 25s) Loss: 1.1485 Grad: 3.1235


Train: 100%|████████████████████████████| 26/26 [11:24<00:00, 26.34s/train_ba


Epoch: [4][25/26] Elapsed 11m 24s (remain 0m 0s) Loss: 1.1658 Grad: 3.3045
train_epoch took 684.7745 seconds.


Validation:  11%|███                        | 1/9 [00:09<01:16,  9.59s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.3904


Validation: 100%|████████████████████████████| 9/9 [01:37<00:00, 10.78s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.2071
valid_epoch took 97.0378 seconds.
Epoch 4 - avg_train_loss: 1.1658  avg_val_loss: 1.2071  time: 782s


Train:   4%|█                            | 1/26 [00:58<24:14, 58.20s/train_ba

Epoch: [5][0/26] Elapsed 0m 58s (remain 24m 14s) Loss: 1.1596 Grad: 2.3584


Train:  81%|███████████████████▍     | 21/26 [1:04:57<06:16, 75.34s/train_ba

Epoch: [5][20/26] Elapsed 64m 57s (remain 15m 28s) Loss: 1.0601 Grad: 2.4875


Train: 100%|████████████████████████████| 26/26 [1:06:03<00:00, 152.46s/train_ba


Epoch: [5][25/26] Elapsed 66m 3s (remain 0m 0s) Loss: 1.0262 Grad: 2.5446  L
train_epoch took 3963.8953 seconds.
```

```
Validation:  11%|██                    | 1/9 [00:09<01:17,  9.66s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 17s) Loss: 1.2818

Validation: 100%|████████████████████| 9/9 [01:39<00:00, 11.00s/valid_ba

EVAL: [8/9] Elapsed 1m 39s (remain 0m 0s) Loss: 1.1214
valid_epoch took 99.0202 seconds.
Epoch 5 - avg_train_loss: 1.0262  avg_val_loss: 1.1214  time: 4063s

Train:   4%|▊                       | 1/26 [00:08<03:43,  8.92s/train_ba

Epoch: [6][0/26] Elapsed 0m 8s (remain 3m 43s) Loss: 1.1130 Grad: 3.3631  LR

Train:  81%|████████████████▏     | 21/26 [10:38<02:05, 25.04s/train_ba

Epoch: [6][20/26] Elapsed 10m 38s (remain 2m 32s) Loss: 0.9991 Grad: 4.7167

Train: 100%|████████████████████| 26/26 [12:48<00:00, 29.56s/train_ba

Epoch: [6][25/26] Elapsed 12m 48s (remain 0m 0s) Loss: 0.9518 Grad: 2.7989
train_epoch took 768.6919 seconds.

Validation:  11%|██                    | 1/9 [00:09<01:16,  9.60s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.2197

Validation: 100%|████████████████████| 9/9 [01:37<00:00, 10.78s/valid_ba
```

```
EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.1211
valid_epoch took 97.0250 seconds.
Epoch 6 - avg_train_loss: 0.9518  avg_val_loss: 1.1211  time: 866s


Train:   4%|▉                          | 1/26 [00:09<03:47,  9.12s/train_ba

Epoch: [7][0/26] Elapsed 0m 9s (remain 3m 47s) Loss: 0.9140 Grad: 2.6752  LR


Train:  81%|████████████████████     | 21/26 [11:18<01:38, 19.61s/train_ba

Epoch: [7][20/26] Elapsed 11m 18s (remain 2m 41s) Loss: 0.8836 Grad: 3.9744


Train: 100%|█████████████████████████| 26/26 [12:23<00:00, 28.61s/train_ba


Epoch: [7][25/26] Elapsed 12m 23s (remain 0m 0s) Loss: 0.8747 Grad: 2.8473
train_epoch took 743.9821 seconds.


Validation:  11%|██                      | 1/9 [00:09<01:16,  9.59s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.1816


Validation: 100%|████████████████████████| 9/9 [01:37<00:00, 10.85s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.0561
valid_epoch took 97.6387 seconds.
Epoch 7 - avg_train_loss: 0.8747  avg_val_loss: 1.0561  time: 842s


Train:   4%|▉                          | 1/26 [00:09<04:02,  9.71s/train_ba

Epoch: [8][0/26] Elapsed 0m 9s (remain 4m 2s) Loss: 0.9529 Grad: 3.2725  LR:
```

```
Train:  81%|████████████████████     | 21/26 [11:55<03:12, 38.49s/train_ba

Epoch: [8][20/26] Elapsed 11m 55s (remain 2m 50s) Loss: 0.8344 Grad: 6.0647


Train: 100%|█████████████████████████| 26/26 [12:49<00:00, 29.60s/train_ba

Epoch: [8][25/26] Elapsed 12m 49s (remain 0m 0s) Loss: 0.7984 Grad: 3.7328
train_epoch took 769.4999 seconds.


Validation:  11%|██                        | 1/9 [00:09<01:16,  9.58s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.2122


Validation: 100%|█████████████████████████| 9/9 [01:37<00:00, 10.79s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.0373
valid_epoch took 97.1049 seconds.
Epoch 8 - avg_train_loss: 0.7984  avg_val_loss: 1.0373  time: 867s


Train:   4%|█                           | 1/26 [00:32<13:39, 32.76s/train_ba

Epoch: [9][0/26] Elapsed 0m 32s (remain 13m 39s) Loss: 0.6527 Grad: 2.9504


Train:  81%|████████████████████     | 21/26 [10:44<01:50, 22.09s/train_ba

Epoch: [9][20/26] Elapsed 10m 44s (remain 2m 33s) Loss: 0.7421 Grad: 3.6401


Train: 100%|█████████████████████████| 26/26 [12:20<00:00, 28.47s/train_ba
```

```
Epoch: [9][25/26] Elapsed 12m 20s (remain 0m 0s) Loss: 0.7326 Grad: 3.2860
train_epoch took 740.2389 seconds.


Validation:  11%|██                    | 1/9 [00:09<01:18,  9.87s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 18s) Loss: 1.2361


Validation: 100%|████████████████████| 9/9 [01:37<00:00, 10.86s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.0585
valid_epoch took 97.7077 seconds.
Epoch 9 - avg_train_loss: 0.7326  avg_val_loss: 1.0585  time: 838s


Train:   4%|▊                         | 1/26 [00:07<03:06,  7.44s/train_ba

Epoch: [10][0/26] Elapsed 0m 7s (remain 3m 6s) Loss: 0.7470 Grad: 6.0681  LR


Train:  81%|████████████████▏    | 21/26 [09:14<01:37, 19.54s/train_ba

Epoch: [10][20/26] Elapsed 9m 14s (remain 2m 12s) Loss: 0.7305 Grad: 3.8500


Train: 100%|████████████████████| 26/26 [12:43<00:00, 29.38s/train_ba


Epoch: [10][25/26] Elapsed 12m 43s (remain 0m 0s) Loss: 0.7212 Grad: 4.6630
train_epoch took 763.9669 seconds.


Validation:  11%|██                    | 1/9 [00:09<01:16,  9.50s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.1101
```

```
Validation: 100%|████████████████████████| 9/9 [01:37<00:00, 10.80s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.0286
valid_epoch took 97.2178 seconds.
Epoch 10 - avg_train_loss: 0.7212  avg_val_loss: 1.0286  time: 861s


Train:   4%|▉                          | 1/26 [00:09<04:03,  9.74s/train_ba

Epoch: [11][0/26] Elapsed 0m 9s (remain 4m 3s) Loss: 0.6824 Grad: 4.5213  LR


Train:  81%|████████████████████     | 21/26 [11:22<01:41, 20.27s/train_ba

Epoch: [11][20/26] Elapsed 11m 22s (remain 2m 42s) Loss: 0.6725 Grad: 4.0601


Train: 100%|████████████████████████| 26/26 [12:36<00:00, 29.10s/train_ba


Epoch: [11][25/26] Elapsed 12m 36s (remain 0m 0s) Loss: 0.6996 Grad: 4.9381
train_epoch took 756.5016 seconds.


Validation:  11%|██                        | 1/9 [00:09<01:16,  9.58s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.0706


Validation: 100%|████████████████████████| 9/9 [01:38<00:00, 10.97s/valid_ba


EVAL: [8/9] Elapsed 1m 38s (remain 0m 0s) Loss: 0.9728
valid_epoch took 98.7442 seconds.
Epoch 11 - avg_train_loss: 0.6996  avg_val_loss: 0.9728  time: 855s
```

```
Train:   4%|▉                              | 1/26 [00:13<05:36, 13.46s/train_ba

Epoch: [12][0/26] Elapsed 0m 13s (remain 5m 36s) Loss: 0.6056 Grad: 3.7459


Train:  81%|████████████████████▎     | 21/26 [11:23<03:09, 38.00s/train_ba

Epoch: [12][20/26] Elapsed 11m 23s (remain 2m 42s) Loss: 0.6047 Grad: 2.5716


Train: 100%|█████████████████████████▉| 26/26 [12:44<00:00, 29.39s/train_ba


Epoch: [12][25/26] Elapsed 12m 44s (remain 0m 0s) Loss: 0.5958 Grad: 3.8759
train_epoch took 764.0264 seconds.


Validation:  11%|██▊                        | 1/9 [00:09<01:16,  9.54s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.1181


Validation: 100%|████████████████████████| 9/9 [01:36<00:00, 10.76s/valid_ba


EVAL: [8/9] Elapsed 1m 36s (remain 0m 0s) Loss: 0.9888
valid_epoch took 96.8539 seconds.
Epoch 12 - avg_train_loss: 0.5958  avg_val_loss: 0.9888  time: 861s


Train:   4%|▉                              | 1/26 [00:14<05:57, 14.30s/train_ba

Epoch: [13][0/26] Elapsed 0m 14s (remain 5m 57s) Loss: 0.6161 Grad: 4.2520


Train:  81%|████████████████████▎     | 21/26 [07:37<00:49,  9.83s/train_ba

Epoch: [13][20/26] Elapsed 7m 37s (remain 1m 48s) Loss: 0.5918 Grad: 2.9083
```

```
Train: 100%|████████████████████████████| 26/26 [12:44<00:00, 29.40s/train_ba


Epoch: [13][25/26] Elapsed 12m 44s (remain 0m 0s) Loss: 0.6151 Grad: 3.4268
train_epoch took 764.3253 seconds.


Validation:  11%|███                        | 1/9 [00:09<01:15,  9.49s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 15s) Loss: 1.1218


Validation: 100%|████████████████████████████| 9/9 [01:38<00:00, 10.97s/valid_ba


EVAL: [8/9] Elapsed 1m 38s (remain 0m 0s) Loss: 0.9935
valid_epoch took 98.7773 seconds.
Epoch 13 - avg_train_loss: 0.6151  avg_val_loss: 0.9935  time: 863s


Train:   4%|█                           | 1/26 [03:08<1:18:30, 188.42s/train_ba

Epoch: [14][0/26] Elapsed 3m 8s (remain 78m 30s) Loss: 0.5323 Grad: 3.7330


Train:  81%|███████████████████████     | 21/26 [09:52<01:25, 17.16s/train_ba

Epoch: [14][20/26] Elapsed 9m 52s (remain 2m 20s) Loss: 0.6612 Grad: 3.5700


Train: 100%|████████████████████████████| 26/26 [12:48<00:00, 29.55s/train_ba


Epoch: [14][25/26] Elapsed 12m 48s (remain 0m 0s) Loss: 0.6178 Grad: 3.5023
train_epoch took 768.4253 seconds.


Validation:  11%|███                        | 1/9 [00:09<01:15,  9.42s/valid_ba
```

```
EVAL: [0/9] Elapsed 0m 9s (remain 1m 15s) Loss: 1.0635


Validation: 100%|████████████████████████| 9/9 [01:38<00:00, 11.00s/valid_ba

EVAL: [8/9] Elapsed 1m 38s (remain 0m 0s) Loss: 0.9859
valid_epoch took 98.9809 seconds.
Epoch 14 - avg_train_loss: 0.6178  avg_val_loss: 0.9859  time: 867s
Early stopping triggered at 13 epochs without improvement.




train_loop took 15774.7488 seconds.
get_result took 0.0028 seconds.
get_result took 0.0004 seconds.
========== Fold 0 Stage 1 result: 1.2593094134672695 ==========
Starting Stage 2 Training for Fold 0
Training Stage 2: Filtering data based on KL Loss < 9


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
  warnings.warn(
Train:   0%|                                        | 0/25 [00:00<?, ?train_batc
  warnings.warn(
Train:   4%|█                           | 1/25 [00:11<04:25, 11.07s/train_ba

Epoch: [1][0/25] Elapsed 0m 11s (remain 4m 25s) Loss: 0.3597 Grad: 2.5711  L


Train:  84%|████████████████████▉     | 21/25 [09:57<01:08, 17.22s/train_ba

Epoch: [1][20/25] Elapsed 9m 57s (remain 1m 53s) Loss: 0.6011 Grad: 5.4451


Train: 100%|████████████████████████████| 25/25 [12:32<00:00, 30.08s/train_ba
```

```
Epoch: [1][24/25] Elapsed 12m 32s (remain 0m 0s) Loss: 0.5971 Grad: 4.1627
train_epoch took 752.1283 seconds.


Validation:  11%|██                    | 1/9 [00:09<01:19,  9.91s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 19s) Loss: 1.0992


Validation: 100%|████████████████████| 9/9 [01:38<00:00, 10.97s/valid_ba


EVAL: [8/9] Elapsed 1m 38s (remain 0m 0s) Loss: 1.0065
valid_epoch took 98.6980 seconds.
Epoch 1 - avg_train_loss: 0.5971  avg_val_loss: 1.0065  time: 851s


Train:   4%|▊                         | 1/25 [00:11<04:35, 11.48s/train_ba

Epoch: [2][0/25] Elapsed 0m 11s (remain 4m 35s) Loss: 0.7057 Grad: 4.8587  L


Train:  84%|████████████████▊    | 21/25 [08:19<01:46, 26.65s/train_ba

Epoch: [2][20/25] Elapsed 8m 19s (remain 1m 35s) Loss: 0.5661 Grad: 4.0058


Train: 100%|████████████████████| 25/25 [11:55<00:00, 28.60s/train_ba


Epoch: [2][24/25] Elapsed 11m 55s (remain 0m 0s) Loss: 0.5640 Grad: 6.0272
train_epoch took 715.0197 seconds.


Validation:  11%|██                    | 1/9 [00:09<01:15,  9.46s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 15s) Loss: 1.1733
```

```
Validation: 100%|████████████████████████| 9/9 [01:36<00:00, 10.71s/valid_ba


EVAL: [8/9] Elapsed 1m 36s (remain 0m 0s) Loss: 1.0389
valid_epoch took 96.3735 seconds.
Epoch 2 - avg_train_loss: 0.5640  avg_val_loss: 1.0389  time: 811s


Train:   4%|▉                         | 1/25 [00:09<03:36,  9.01s/train_ba

Epoch: [3][0/25] Elapsed 0m 9s (remain 3m 36s) Loss: 0.7322 Grad: 5.7345  LR


Train:  84%|████████████████████▉    | 21/25 [10:08<02:38, 39.62s/train_ba

Epoch: [3][20/25] Elapsed 10m 8s (remain 1m 55s) Loss: 0.5466 Grad: 3.5721


Train: 100%|████████████████████████| 25/25 [12:14<00:00, 29.38s/train_ba


Epoch: [3][24/25] Elapsed 12m 14s (remain 0m 0s) Loss: 0.5382 Grad: 5.5361
train_epoch took 734.4022 seconds.


Validation:  11%|██▊                      | 1/9 [00:09<01:16,  9.55s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.1754


Validation: 100%|████████████████████████| 9/9 [01:37<00:00, 10.81s/valid_ba


EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.1076
valid_epoch took 97.3329 seconds.
Epoch 3 - avg_train_loss: 0.5382  avg_val_loss: 1.1076  time: 832s
```

```
Train:   4%|▉                              | 1/25 [00:09<03:51,  9.66s/train_ba

Epoch: [4][0/25] Elapsed 0m 9s (remain 3m 51s) Loss: 0.6783 Grad: 7.5791  LR


Train:  84%|████████████████████▉    | 21/25 [10:26<01:50, 27.67s/train_ba

Epoch: [4][20/25] Elapsed 10m 26s (remain 1m 59s) Loss: 0.4406 Grad: 3.4982


Train: 100%|█████████████████████████| 25/25 [12:09<00:00, 29.18s/train_ba


Epoch: [4][24/25] Elapsed 12m 9s (remain 0m 0s) Loss: 0.4421 Grad: 5.6365  L
train_epoch took 729.4974 seconds.


Validation:  11%|██▊                      | 1/9 [00:09<01:16,  9.57s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 16s) Loss: 1.0717


Validation: 100%|█████████████████████████| 9/9 [01:36<00:00, 10.78s/valid_ba


EVAL: [8/9] Elapsed 1m 36s (remain 0m 0s) Loss: 1.1296
valid_epoch took 96.9878 seconds.
Epoch 4 - avg_train_loss: 0.4421  avg_val_loss: 1.1296  time: 826s


Train:   4%|▉                              | 1/25 [00:09<03:51,  9.65s/train_ba

Epoch: [5][0/25] Elapsed 0m 9s (remain 3m 51s) Loss: 0.3917 Grad: 4.1032  LR


Train:  84%|████████████████████▉    | 21/25 [55:03<40:38, 609.68s/train_ba

Epoch: [5][20/25] Elapsed 55m 3s (remain 10m 29s) Loss: 0.4908 Grad: 12.6854
```

```
Train: 100%|████████████████████████| 25/25 [56:07<00:00, 134.68s/train_ba


Epoch: [5][24/25] Elapsed 56m 7s (remain 0m 0s) Loss: 0.4707 Grad: 4.6297  L
train_epoch took 3367.0285 seconds.


Validation:  11%|██                      | 1/9 [00:09<01:19,  9.89s/valid_ba

EVAL: [0/9] Elapsed 0m 9s (remain 1m 19s) Loss: 1.1963


Validation: 100%|████████████████████████| 9/9 [01:37<00:00, 10.80s/valid_ba

EVAL: [8/9] Elapsed 1m 37s (remain 0m 0s) Loss: 1.0837
valid_epoch took 97.2333 seconds.
Epoch 5 - avg_train_loss: 0.4707  avg_val_loss: 1.0837  time: 3464s
Early stopping triggered at 4 epochs without improvement.




train_loop took 6785.7628 seconds.
get_result took 0.0016 seconds.
get_result took 0.0003 seconds.
========== Fold 0 Stage 2 result: 1.2543410266495378 ==========
Starting Stage 1 Training for Fold 1
Training Stage 1: Using all data


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
  warnings.warn(
Train:   0%|                                  | 0/26 [00:00<?, ?train_batc
  warnings.warn(
Train:   4%|█                           | 1/26 [00:08<03:39,  8.76s/train_ba

Epoch: [1][0/26] Elapsed 0m 8s (remain 3m 39s) Loss: 1.2886 Grad: 2.0540  LR
```

```
Train:  81%|████████████████████     | 21/26 [07:01<01:13, 14.63s/train_ba

Epoch: [1][20/26] Elapsed 7m 1s (remain 1m 40s) Loss: 1.3886 Grad: 2.1359  L

Train: 100%|████████████████████████| 26/26 [11:22<00:00, 26.25s/train_ba

Epoch: [1][25/26] Elapsed 11m 22s (remain 0m 0s) Loss: 1.3965 Grad: 2.7388
train_epoch took 682.5236 seconds.

Validation:  11%|██                       | 1/9 [01:31<12:11, 91.49s/valid_ba

EVAL: [0/9] Elapsed 1m 31s (remain 12m 11s) Loss: 1.0645

Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.42s/valid_ba

EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.4295
valid_epoch took 237.7739 seconds.
Epoch 1 - avg_train_loss: 1.3965  avg_val_loss: 1.4295  time: 920s

Train:   4%|▉                          | 1/26 [00:49<20:34, 49.40s/train_ba

Epoch: [2][0/26] Elapsed 0m 49s (remain 20m 34s) Loss: 1.5393 Grad: 2.7127

Train:  81%|████████████████████     | 21/26 [05:47<01:22, 16.57s/train_ba

Epoch: [2][20/26] Elapsed 5m 47s (remain 1m 22s) Loss: 1.3349 Grad: 2.1129

Train: 100%|████████████████████████| 26/26 [10:21<00:00, 23.91s/train_ba
```

```
Epoch: [2][25/26] Elapsed 10m 21s (remain 0m 0s) Loss: 1.3411 Grad: 2.8941
train_epoch took 621.6001 seconds.


Validation:  11%|██                    | 1/9 [01:30<12:04, 90.51s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 4s) Loss: 1.0792


Validation: 100%|████████████████████| 9/9 [03:52<00:00, 25.82s/valid_ba


EVAL: [8/9] Elapsed 3m 52s (remain 0m 0s) Loss: 1.4168
valid_epoch took 232.4094 seconds.
Epoch 2 - avg_train_loss: 1.3411  avg_val_loss: 1.4168  time: 854s


Train:   4%|▊                      | 1/26 [00:16<07:00, 16.82s/train_ba

Epoch: [3][0/26] Elapsed 0m 16s (remain 7m 0s) Loss: 1.4311 Grad: 4.2622  LR


Train:  81%|████████████████▎    | 21/26 [06:41<01:02, 12.41s/train_ba

Epoch: [3][20/26] Elapsed 6m 41s (remain 1m 35s) Loss: 1.2560 Grad: 5.6275


Train: 100%|████████████████████| 26/26 [10:24<00:00, 24.00s/train_ba


Epoch: [3][25/26] Elapsed 10m 24s (remain 0m 0s) Loss: 1.2353 Grad: 2.9050
train_epoch took 624.1211 seconds.


Validation:  11%|██                    | 1/9 [01:30<12:02, 90.37s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 3s) Loss: 0.9895
```

```
Validation: 100%|████████████████████████| 9/9 [03:51<00:00, 25.77s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.3286
valid_epoch took 231.9070 seconds.
Epoch 3 - avg_train_loss: 1.2353  avg_val_loss: 1.3286  time: 856s


Train:   4%|▉                         | 1/26 [00:17<07:26, 17.86s/train_ba

Epoch: [4][0/26] Elapsed 0m 17s (remain 7m 26s) Loss: 1.1795 Grad: 2.4190  L


Train:  81%|████████████████████     | 21/26 [09:30<01:27, 17.46s/train_ba

Epoch: [4][20/26] Elapsed 9m 30s (remain 2m 15s) Loss: 1.1153 Grad: 3.2731


Train: 100%|████████████████████████| 26/26 [10:20<00:00, 23.85s/train_ba


Epoch: [4][25/26] Elapsed 10m 20s (remain 0m 0s) Loss: 1.1238 Grad: 2.8936
train_epoch took 620.1579 seconds.


Validation:  11%|██                       | 1/9 [01:30<12:01, 90.17s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.8902


Validation: 100%|████████████████████████| 9/9 [03:51<00:00, 25.72s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.2076
valid_epoch took 231.5014 seconds.
Epoch 4 - avg_train_loss: 1.1238  avg_val_loss: 1.2076  time: 852s
```

```
Train:   4%|▉                              | 1/26 [00:08<03:26,  8.24s/train_ba

Epoch: [5][0/26] Elapsed 0m 8s (remain 3m 26s) Loss: 0.9940 Grad: 2.6176  LR


Train:  81%|████████████████████▉     | 21/26 [06:56<02:30, 30.14s/train_ba

Epoch: [5][20/26] Elapsed 6m 56s (remain 1m 39s) Loss: 1.0533 Grad: 2.7355


Train: 100%|█████████████████████████▉| 26/26 [10:23<00:00, 23.98s/train_ba


Epoch: [5][25/26] Elapsed 10m 23s (remain 0m 0s) Loss: 1.0338 Grad: 3.5511
train_epoch took 623.5845 seconds.


Validation:  11%|██▊                        | 1/9 [01:30<12:01, 90.18s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.8770


Validation: 100%|████████████████████████| 9/9 [03:51<00:00, 25.72s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.1680
valid_epoch took 231.5265 seconds.
Epoch 5 - avg_train_loss: 1.0338  avg_val_loss: 1.1680  time: 855s


Train:   4%|▉                              | 1/26 [00:22<09:32, 22.90s/train_ba

Epoch: [6][0/26] Elapsed 0m 22s (remain 9m 32s) Loss: 0.8670 Grad: 2.8064  L


Train:  81%|████████████████████▉     | 21/26 [09:20<01:10, 14.06s/train_ba

Epoch: [6][20/26] Elapsed 9m 20s (remain 2m 13s) Loss: 0.9402 Grad: 3.4267
```

```
Train: 100%|████████████████████████████| 26/26 [10:24<00:00, 24.00s/train_ba


Epoch: [6][25/26] Elapsed 10m 24s (remain 0m 0s) Loss: 0.9184 Grad: 4.0653
train_epoch took 624.1122 seconds.


Validation:  11%|███                        | 1/9 [01:30<12:04, 90.52s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 4s) Loss: 0.8676


Validation: 100%|████████████████████████████| 9/9 [03:52<00:00, 25.83s/valid_ba


EVAL: [8/9] Elapsed 3m 52s (remain 0m 0s) Loss: 1.1018
valid_epoch took 232.4863 seconds.
Epoch 6 - avg_train_loss: 0.9184  avg_val_loss: 1.1018  time: 857s


Train:   4%|█                            | 1/26 [00:08<03:23,  8.15s/train_ba

Epoch: [7][0/26] Elapsed 0m 8s (remain 3m 23s) Loss: 1.3200 Grad: 3.5590  LR


Train:  81%|███████████████████████      | 21/26 [06:40<02:43, 32.65s/train_ba

Epoch: [7][20/26] Elapsed 6m 40s (remain 1m 35s) Loss: 0.9000 Grad: 3.4403


Train: 100%|████████████████████████████| 26/26 [10:24<00:00, 24.01s/train_ba


Epoch: [7][25/26] Elapsed 10m 24s (remain 0m 0s) Loss: 0.8688 Grad: 5.1446
train_epoch took 624.2930 seconds.


Validation:  11%|███                        | 1/9 [01:30<12:02, 90.33s/valid_ba
```

```
EVAL: [0/9] Elapsed 1m 30s (remain 12m 2s) Loss: 0.8520


Validation: 100%|████████████████████████| 9/9 [03:52<00:00, 25.78s/valid_ba


EVAL: [8/9] Elapsed 3m 52s (remain 0m 0s) Loss: 1.0657
valid_epoch took 232.0278 seconds.
Epoch 7 - avg_train_loss: 0.8688  avg_val_loss: 1.0657  time: 856s


Train:   4%|▉                         | 1/26 [00:07<03:17,  7.92s/train_ba

Epoch: [8][0/26] Elapsed 0m 7s (remain 3m 18s) Loss: 0.8885 Grad: 2.9915  LR


Train:  81%|████████████████████     | 21/26 [07:44<01:55, 23.06s/train_ba

Epoch: [8][20/26] Elapsed 7m 44s (remain 1m 50s) Loss: 0.7812 Grad: 3.8306


Train: 100%|████████████████████████| 26/26 [09:34<00:00, 22.09s/train_ba


Epoch: [8][25/26] Elapsed 9m 34s (remain 0m 0s) Loss: 0.7827 Grad: 3.6504  L
train_epoch took 574.3482 seconds.


Validation:  11%|██                       | 1/9 [01:30<12:01, 90.13s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.8127


Validation: 100%|████████████████████████| 9/9 [03:52<00:00, 25.78s/valid_ba


EVAL: [8/9] Elapsed 3m 52s (remain 0m 0s) Loss: 1.0651
valid_epoch took 232.0536 seconds.
```

```
Epoch 8 - avg_train_loss: 0.7827  avg_val_loss: 1.0651  time: 806s


Train:   4%|▉                          | 1/26 [00:15<06:33, 15.75s/train_ba

Epoch: [9][0/26] Elapsed 0m 15s (remain 6m 33s) Loss: 0.6626 Grad: 3.6092  L


Train:  81%|████████████████████     | 21/26 [07:07<02:48, 33.68s/train_ba

Epoch: [9][20/26] Elapsed 7m 7s (remain 1m 41s) Loss: 0.7527 Grad: 4.7718  L


Train: 100%|█████████████████████████| 26/26 [10:19<00:00, 23.84s/train_ba


Epoch: [9][25/26] Elapsed 10m 19s (remain 0m 0s) Loss: 0.7559 Grad: 3.2724
train_epoch took 619.7162 seconds.


Validation:  11%|██                        | 1/9 [01:30<12:02, 90.31s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 2s) Loss: 0.7528


Validation: 100%|█████████████████████████| 9/9 [03:51<00:00, 25.75s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0442
valid_epoch took 231.7170 seconds.
Epoch 9 - avg_train_loss: 0.7559  avg_val_loss: 1.0442  time: 851s


Train:   4%|▉                          | 1/26 [00:08<03:36,  8.68s/train_ba

Epoch: [10][0/26] Elapsed 0m 8s (remain 3m 36s) Loss: 0.6011 Grad: 3.4805  L


Train:  81%|████████████████████     | 21/26 [06:50<01:35, 19.14s/train_ba
```

```
Epoch: [10][20/26] Elapsed 6m 50s (remain 1m 37s) Loss: 0.6271 Grad: 2.8379

Train: 100%|████████████████████████████| 26/26 [10:21<00:00, 23.89s/train_ba

Epoch: [10][25/26] Elapsed 10m 21s (remain 0m 0s) Loss: 0.6443 Grad: 4.1648
train_epoch took 621.1984 seconds.

Validation:  11%|███                        | 1/9 [01:30<12:02, 90.37s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 2s) Loss: 0.7550

Validation: 100%|████████████████████████████| 9/9 [03:51<00:00, 25.75s/valid_ba

EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0457
valid_epoch took 231.7687 seconds.
Epoch 10 - avg_train_loss: 0.6443  avg_val_loss: 1.0457  time: 853s

Train:   4%|█                            | 1/26 [00:28<11:48, 28.32s/train_ba

Epoch: [11][0/26] Elapsed 0m 28s (remain 11m 48s) Loss: 0.8225 Grad: 4.8326

Train:  81%|███████████████████████      | 21/26 [09:15<01:38, 19.65s/train_ba

Epoch: [11][20/26] Elapsed 9m 15s (remain 2m 12s) Loss: 0.6774 Grad: 5.0555

Train: 100%|████████████████████████████| 26/26 [10:21<00:00, 23.90s/train_ba

Epoch: [11][25/26] Elapsed 10m 21s (remain 0m 0s) Loss: 0.6533 Grad: 4.4661
train_epoch took 621.3444 seconds.
```

```
Validation:  11%|██                    | 1/9 [01:30<12:01, 90.20s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7293

Validation: 100%|████████████████████| 9/9 [03:51<00:00, 25.73s/valid_ba

EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0459
valid_epoch took 231.5402 seconds.
Epoch 11 - avg_train_loss: 0.6533  avg_val_loss: 1.0459  time: 853s

Train:   4%|▊                       | 1/26 [00:10<04:34, 10.97s/train_ba

Epoch: [12][0/26] Elapsed 0m 10s (remain 4m 34s) Loss: 0.7490 Grad: 4.4755

Train:  81%|████████████████▏     | 21/26 [07:14<01:06, 13.33s/train_ba

Epoch: [12][20/26] Elapsed 7m 14s (remain 1m 43s) Loss: 0.6395 Grad: 4.0626

Train: 100%|████████████████████| 26/26 [10:20<00:00, 23.85s/train_ba

Epoch: [12][25/26] Elapsed 10m 20s (remain 0m 0s) Loss: 0.6253 Grad: 3.7060
train_epoch took 620.1603 seconds.

Validation:  11%|██                    | 1/9 [01:30<12:01, 90.21s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7828

Validation: 100%|████████████████████| 9/9 [03:51<00:00, 25.73s/valid_ba
```

```
EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0371
valid_epoch took 231.5905 seconds.
Epoch 12 - avg_train_loss: 0.6253  avg_val_loss: 1.0371  time: 852s


Train:   4%|▉                          | 1/26 [00:08<03:33,  8.53s/train_ba

Epoch: [13][0/26] Elapsed 0m 8s (remain 3m 33s) Loss: 0.5575 Grad: 3.7055  L


Train:  81%|████████████████████     | 21/26 [08:33<01:10, 14.16s/train_ba

Epoch: [13][20/26] Elapsed 8m 33s (remain 2m 2s) Loss: 0.6092 Grad: 5.2785


Train: 100%|█████████████████████████| 26/26 [10:20<00:00, 23.88s/train_ba


Epoch: [13][25/26] Elapsed 10m 20s (remain 0m 0s) Loss: 0.6178 Grad: 3.3107
train_epoch took 620.9670 seconds.


Validation:  11%|██▊                      | 1/9 [01:30<12:01, 90.16s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7683


Validation: 100%|█████████████████████████| 9/9 [03:51<00:00, 25.74s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0458
valid_epoch took 231.6988 seconds.
Epoch 13 - avg_train_loss: 0.6178  avg_val_loss: 1.0458  time: 853s


Train:   4%|▉                          | 1/26 [00:08<03:36,  8.67s/train_ba

Epoch: [14][0/26] Elapsed 0m 8s (remain 3m 36s) Loss: 0.6671 Grad: 4.3845  L
```

```
Train:  81%|████████████████████     | 21/26 [08:21<04:47, 57.47s/train_ba

Epoch: [14][20/26] Elapsed 8m 21s (remain 1m 59s) Loss: 0.5594 Grad: 2.6566


Train: 100%|████████████████████████| 26/26 [10:18<00:00, 23.77s/train_ba


Epoch: [14][25/26] Elapsed 10m 18s (remain 0m 0s) Loss: 0.5675 Grad: 3.5370
train_epoch took 618.0848 seconds.


Validation:  11%|██                       | 1/9 [01:30<12:01, 90.15s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7892


Validation: 100%|████████████████████████| 9/9 [03:51<00:00, 25.73s/valid_ba

EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0466
valid_epoch took 231.5667 seconds.
Epoch 14 - avg_train_loss: 0.5675  avg_val_loss: 1.0466  time: 850s
Early stopping triggered at 13 epochs without improvement.




train_loop took 11969.1596 seconds.
get_result took 0.0018 seconds.
get_result took 0.0004 seconds.
========== Fold 1 Stage 1 result: 1.2588057879216985 ==========
Starting Stage 2 Training for Fold 1
Training Stage 2: Filtering data based on KL Loss < 9


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
```

```
  warnings.warn(
Train:   0%|                                        | 0/25 [00:00<?, ?train_batc
  warnings.warn(
Train:   4%|▉                                | 1/25 [00:09<03:52,  9.67s/train_ba

Epoch: [1][0/25] Elapsed 0m 9s (remain 3m 52s) Loss: 0.5492 Grad: 2.9689  LR


Train:  84%|████████████████████▉    | 21/25 [08:08<01:04, 16.22s/train_ba

Epoch: [1][20/25] Elapsed 8m 8s (remain 1m 33s) Loss: 0.6193 Grad: 4.8152  L


Train: 100%|█████████████████████████| 25/25 [09:48<00:00, 23.52s/train_ba


Epoch: [1][24/25] Elapsed 9m 48s (remain 0m 0s) Loss: 0.6024 Grad: 3.8268  L
train_epoch took 588.0198 seconds.


Validation:  11%|██▊                      | 1/9 [01:30<12:01, 90.15s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7526


Validation: 100%|█████████████████████████| 9/9 [03:51<00:00, 25.73s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0412
valid_epoch took 231.5844 seconds.
Epoch 1 - avg_train_loss: 0.6024  avg_val_loss: 1.0412  time: 820s


Train:   4%|▉                                | 1/25 [00:08<03:15,  8.16s/train_ba

Epoch: [2][0/25] Elapsed 0m 8s (remain 3m 15s) Loss: 0.3977 Grad: 3.2281  LR


Train:  84%|████████████████████▉    | 21/25 [06:33<01:27, 21.99s/train_ba
```

```
Epoch: [2][20/25] Elapsed 6m 33s (remain 1m 14s) Loss: 0.5272 Grad: 3.8525


Train: 100%|████████████████████████| 25/25 [09:48<00:00, 23.52s/train_ba


Epoch: [2][24/25] Elapsed 9m 48s (remain 0m 0s) Loss: 0.5259 Grad: 4.4489  L
train_epoch took 588.0533 seconds.


Validation:  11%|██                     | 1/9 [01:30<12:01, 90.20s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7166


Validation: 100%|████████████████████████| 9/9 [03:51<00:00, 25.69s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0335
valid_epoch took 231.1923 seconds.
Epoch 2 - avg_train_loss: 0.5259  avg_val_loss: 1.0335  time: 819s


Train:   4%|▉                         | 1/25 [00:25<10:12, 25.54s/train_ba

Epoch: [3][0/25] Elapsed 0m 25s (remain 10m 12s) Loss: 0.4761 Grad: 3.9144


Train:  84%|████████████████████    | 21/25 [08:58<02:49, 42.40s/train_ba

Epoch: [3][20/25] Elapsed 8m 58s (remain 1m 42s) Loss: 0.5233 Grad: 3.8256


Train: 100%|████████████████████████| 25/25 [09:49<00:00, 23.56s/train_ba


Epoch: [3][24/25] Elapsed 9m 49s (remain 0m 0s) Loss: 0.5339 Grad: 4.2939  L
train_epoch took 589.0142 seconds.
```

```
Validation:  11%|██                    | 1/9 [01:30<12:01, 90.23s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.8092

Validation: 100%|████████████████████| 9/9 [03:51<00:00, 25.73s/valid_ba

EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0532
valid_epoch took 231.5321 seconds.
Epoch 3 - avg_train_loss: 0.5339  avg_val_loss: 1.0532  time: 821s

Train:   4%|▊                      | 1/25 [02:28<59:17, 148.23s/train_ba

Epoch: [4][0/25] Elapsed 2m 28s (remain 59m 17s) Loss: 0.4204 Grad: 3.8256

Train:  84%|████████████████▊    | 21/25 [08:16<01:01, 15.30s/train_ba

Epoch: [4][20/25] Elapsed 8m 16s (remain 1m 34s) Loss: 0.4826 Grad: 4.9745

Train: 100%|████████████████████| 25/25 [09:48<00:00, 23.52s/train_ba

Epoch: [4][24/25] Elapsed 9m 48s (remain 0m 0s) Loss: 0.4993 Grad: 6.5906  L
train_epoch took 588.0075 seconds.

Validation:  11%|██                    | 1/9 [01:30<12:00, 90.12s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 0s) Loss: 0.8359

Validation: 100%|████████████████████| 9/9 [03:51<00:00, 25.69s/valid_ba
```

```
EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.1009
valid_epoch took 231.1736 seconds.
Epoch 4 - avg_train_loss: 0.4993  avg_val_loss: 1.1009  time: 819s


Train:   4%|▉                              | 1/25 [00:08<03:16,  8.19s/train_ba

Epoch: [5][0/25] Elapsed 0m 8s (remain 3m 16s) Loss: 0.2668 Grad: 1.9495  LR


Train:  84%|████████████████████▉     | 21/25 [09:05<01:20, 20.08s/train_ba

Epoch: [5][20/25] Elapsed 9m 5s (remain 1m 43s) Loss: 0.4354 Grad: 3.6939  L


Train: 100%|█████████████████████████| 25/25 [09:48<00:00, 23.53s/train_ba


Epoch: [5][24/25] Elapsed 9m 48s (remain 0m 0s) Loss: 0.4602 Grad: 3.3508  L
train_epoch took 588.1344 seconds.


Validation:  11%|██▊                      | 1/9 [01:30<12:01, 90.17s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7376


Validation: 100%|█████████████████████████| 9/9 [03:51<00:00, 25.71s/valid_ba


EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0734
valid_epoch took 231.4021 seconds.
Epoch 5 - avg_train_loss: 0.4602  avg_val_loss: 1.0734  time: 820s


Train:   4%|▉                              | 1/25 [01:03<25:16, 63.20s/train_ba

Epoch: [6][0/25] Elapsed 1m 3s (remain 25m 16s) Loss: 0.3349 Grad: 4.6870  L
```

```
Train:  84%|████████████████████     | 21/25 [07:24<01:39, 24.96s/train_ba

Epoch: [6][20/25] Elapsed 7m 24s (remain 1m 24s) Loss: 0.4339 Grad: 7.1330


Train: 100%|████████████████████████| 25/25 [08:57<00:00, 21.52s/train_ba


Epoch: [6][24/25] Elapsed 8m 57s (remain 0m 0s) Loss: 0.4321 Grad: 5.0121  L
train_epoch took 537.9590 seconds.


Validation:  11%|██                       | 1/9 [01:30<12:01, 90.13s/valid_ba

EVAL: [0/9] Elapsed 1m 30s (remain 12m 1s) Loss: 0.7554


Validation: 100%|████████████████████████| 9/9 [03:51<00:00, 25.72s/valid_ba

EVAL: [8/9] Elapsed 3m 51s (remain 0m 0s) Loss: 1.0976
valid_epoch took 231.5240 seconds.
Epoch 6 - avg_train_loss: 0.4321  avg_val_loss: 1.0976  time: 769s
Early stopping triggered at 5 epochs without improvement.




train_loop took 4868.6091 seconds.
get_result took 0.0018 seconds.
get_result took 0.0003 seconds.
========== Fold 1 Stage 2 result: 1.251832154335329 ==========
Starting Stage 1 Training for Fold 2
Training Stage 1: Using all data


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
```

```
  warnings.warn(
Train:   0%|                                          | 0/26 [00:00<?, ?train_batc
  warnings.warn(
Train:   4%|▉                                 | 1/26 [00:11<04:58, 11.96s/train_ba

Epoch: [1][0/26] Elapsed 0m 11s (remain 4m 58s) Loss: 1.2622 Grad: 2.6122  L


Train:  81%|████████████████████     | 21/26 [07:49<02:42, 32.58s/train_ba

Epoch: [1][20/26] Elapsed 7m 49s (remain 1m 51s) Loss: 1.4124 Grad: 2.8619


Train: 100%|█████████████████████████| 26/26 [10:14<00:00, 23.65s/train_ba


Epoch: [1][25/26] Elapsed 10m 14s (remain 0m 0s) Loss: 1.4236 Grad: 2.0408
train_epoch took 615.0005 seconds.


Validation:  11%|██                         | 1/9 [00:07<00:56,  7.05s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 1.1946


Validation: 100%|█████████████████████████| 9/9 [03:57<00:00, 26.37s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.2900
valid_epoch took 237.3269 seconds.
Epoch 1 - avg_train_loss: 1.4236  avg_val_loss: 1.2900  time: 852s


Train:   4%|▉                                 | 1/26 [00:08<03:39,  8.79s/train_ba

Epoch: [2][0/26] Elapsed 0m 8s (remain 3m 39s) Loss: 1.4457 Grad: 2.8747  LR


Train:  81%|████████████████████     | 21/26 [07:59<01:35, 19.09s/train_ba
```

```
Epoch: [2][20/26] Elapsed 7m 59s (remain 1m 54s) Loss: 1.3873 Grad: 3.2172


Train: 100%|████████████████████████████| 26/26 [10:15<00:00, 23.68s/train_ba


Epoch: [2][25/26] Elapsed 10m 15s (remain 0m 0s) Loss: 1.3798 Grad: 2.5005
train_epoch took 615.7134 seconds.


Validation:  11%|███                        | 1/9 [00:07<00:56,  7.05s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 1.1604


Validation: 100%|████████████████████████████| 9/9 [03:57<00:00, 26.36s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.2756
valid_epoch took 237.2675 seconds.
Epoch 2 - avg_train_loss: 1.3798  avg_val_loss: 1.2756  time: 853s


Train:   4%|█                            | 1/26 [00:17<07:11, 17.27s/train_ba

Epoch: [3][0/26] Elapsed 0m 17s (remain 7m 11s) Loss: 1.3532 Grad: 3.4627  L


Train:  81%|███████████████████████      | 21/26 [07:10<01:36, 19.28s/train_ba

Epoch: [3][20/26] Elapsed 7m 10s (remain 1m 42s) Loss: 1.2890 Grad: 2.2727


Train: 100%|████████████████████████████| 26/26 [09:24<00:00, 21.70s/train_ba


Epoch: [3][25/26] Elapsed 9m 24s (remain 0m 0s) Loss: 1.2819 Grad: 3.3028  L
train_epoch took 564.1924 seconds.
```

```
Validation:  11%|██                    | 1/9 [00:07<00:56,  7.07s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 1.0740

Validation: 100%|████████████████████| 9/9 [03:57<00:00, 26.39s/valid_ba

EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.2288
valid_epoch took 237.4724 seconds.
Epoch 3 - avg_train_loss: 1.2819  avg_val_loss: 1.2288  time: 802s

Train:   4%|▊                      | 1/26 [00:10<04:12, 10.09s/train_ba

Epoch: [4][0/26] Elapsed 0m 10s (remain 4m 12s) Loss: 1.0507 Grad: 3.0605  L

Train:  81%|████████████████▏    | 21/26 [09:22<01:39, 20.00s/train_ba

Epoch: [4][20/26] Elapsed 9m 22s (remain 2m 13s) Loss: 1.2068 Grad: 3.7690

Train: 100%|████████████████████| 26/26 [10:12<00:00, 23.55s/train_ba

Epoch: [4][25/26] Elapsed 10m 12s (remain 0m 0s) Loss: 1.1839 Grad: 4.0394
train_epoch took 612.2367 seconds.

Validation:  11%|██                    | 1/9 [00:07<00:56,  7.10s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 1.0049

Validation: 100%|████████████████████| 9/9 [03:57<00:00, 26.38s/valid_ba
```

```
EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.1574
valid_epoch took 237.3931 seconds.
Epoch 4 - avg_train_loss: 1.1839  avg_val_loss: 1.1574  time: 850s


Train:   4%|▉                              | 1/26 [00:54<22:42, 54.50s/train_ba

Epoch: [5][0/26] Elapsed 0m 54s (remain 22m 42s) Loss: 1.2269 Grad: 2.6462


Train:  81%|████████████████████     | 21/26 [07:29<01:49, 21.81s/train_ba

Epoch: [5][20/26] Elapsed 7m 29s (remain 1m 46s) Loss: 1.0689 Grad: 2.7762


Train: 100%|████████████████████████| 26/26 [10:03<00:00, 23.23s/train_ba


Epoch: [5][25/26] Elapsed 10m 3s (remain 0m 0s) Loss: 1.0701 Grad: 2.6059  L
train_epoch took 603.8782 seconds.


Validation:  11%|██                      | 1/9 [00:07<00:56,  7.06s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.9530


Validation: 100%|████████████████████| 9/9 [03:57<00:00, 26.36s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.1018
valid_epoch took 237.2417 seconds.
Epoch 5 - avg_train_loss: 1.0701  avg_val_loss: 1.1018  time: 841s


Train:   4%|▉                              | 1/26 [00:20<08:43, 20.93s/train_ba

Epoch: [6][0/26] Elapsed 0m 20s (remain 8m 43s) Loss: 0.9585 Grad: 2.5963  L
```

```
Train:  81%|████████████████████     | 21/26 [07:53<01:55, 23.15s/train_ba

Epoch: [6][20/26] Elapsed 7m 53s (remain 1m 52s) Loss: 1.0099 Grad: 3.0898

Train: 100%|█████████████████████████| 26/26 [10:06<00:00, 23.32s/train_ba

Epoch: [6][25/26] Elapsed 10m 6s (remain 0m 0s) Loss: 0.9769 Grad: 2.3744  L
train_epoch took 606.4171 seconds.

Validation:  11%|██                    | 1/9 [00:07<00:56,  7.11s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.8717

Validation: 100%|█████████████████████| 9/9 [03:57<00:00, 26.38s/valid_ba

EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0519
valid_epoch took 237.4366 seconds.
Epoch 6 - avg_train_loss: 0.9769  avg_val_loss: 1.0519  time: 844s

Train:   4%|▉                         | 1/26 [00:11<04:35, 11.00s/train_ba

Epoch: [7][0/26] Elapsed 0m 11s (remain 4m 35s) Loss: 0.8954 Grad: 3.0160  L

Train:  81%|████████████████████     | 21/26 [09:14<02:00, 24.12s/train_ba

Epoch: [7][20/26] Elapsed 9m 14s (remain 2m 11s) Loss: 0.8876 Grad: 3.0489

Train: 100%|█████████████████████████| 26/26 [10:02<00:00, 23.16s/train_ba
```

```
Epoch: [7][25/26] Elapsed 10m 2s (remain 0m 0s) Loss: 0.8820 Grad: 3.8694  L
train_epoch took 602.1300 seconds.


Validation:  11%|██                    | 1/9 [00:07<00:56,  7.06s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.8333


Validation: 100%|████████████████████| 9/9 [03:57<00:00, 26.34s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0245
valid_epoch took 237.0733 seconds.
Epoch 7 - avg_train_loss: 0.8820  avg_val_loss: 1.0245  time: 839s


Train:   4%|▊                      | 1/26 [00:22<09:10, 22.04s/train_ba

Epoch: [8][0/26] Elapsed 0m 22s (remain 9m 10s) Loss: 1.0306 Grad: 2.6830  L


Train:  81%|████████████████▎    | 21/26 [08:11<03:39, 43.92s/train_ba

Epoch: [8][20/26] Elapsed 8m 11s (remain 1m 56s) Loss: 0.8147 Grad: 2.4032


Train: 100%|████████████████████| 26/26 [10:10<00:00, 23.49s/train_ba


Epoch: [8][25/26] Elapsed 10m 10s (remain 0m 0s) Loss: 0.8184 Grad: 4.2635
train_epoch took 610.6856 seconds.


Validation:  11%|██                    | 1/9 [00:07<00:56,  7.08s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.8173
```

```
Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.35s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0167
valid_epoch took 237.1158 seconds.
Epoch 8 - avg_train_loss: 0.8184  avg_val_loss: 1.0167  time: 848s


Train:   4%|▉                         | 1/26 [00:23<09:49, 23.57s/train_ba

Epoch: [9][0/26] Elapsed 0m 23s (remain 9m 49s) Loss: 0.7365 Grad: 2.9933  L


Train:  81%|████████████████████     | 21/26 [09:03<02:17, 27.58s/train_ba

Epoch: [9][20/26] Elapsed 9m 3s (remain 2m 9s) Loss: 0.7489 Grad: 3.8321  LR


Train: 100%|████████████████████████| 26/26 [10:15<00:00, 23.69s/train_ba


Epoch: [9][25/26] Elapsed 10m 15s (remain 0m 0s) Loss: 0.7713 Grad: 4.3908
train_epoch took 615.9179 seconds.


Validation:  11%|██▊                      | 1/9 [00:07<00:56,  7.08s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7996


Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.36s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0163
valid_epoch took 237.2010 seconds.
Epoch 9 - avg_train_loss: 0.7713  avg_val_loss: 1.0163  time: 853s
```

```
Train:   4%|▉                              | 1/26 [00:07<03:06,  7.46s/train_ba

Epoch: [10][0/26] Elapsed 0m 7s (remain 3m 6s) Loss: 0.7071 Grad: 3.8354  LR


Train:  81%|████████████████████▎     | 21/26 [09:05<01:21, 16.22s/train_ba

Epoch: [10][20/26] Elapsed 9m 5s (remain 2m 9s) Loss: 0.7377 Grad: 3.0709  L


Train: 100%|█████████████████████████▉| 26/26 [10:00<00:00, 23.09s/train_ba


Epoch: [10][25/26] Elapsed 10m 0s (remain 0m 0s) Loss: 0.7312 Grad: 3.2132
train_epoch took 600.4101 seconds.


Validation:  11%|██▊                        | 1/9 [00:07<00:56,  7.07s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.8114


Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.35s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0150
valid_epoch took 237.1938 seconds.
Epoch 10 - avg_train_loss: 0.7312  avg_val_loss: 1.0150  time: 838s


Train:   4%|▉                              | 1/26 [01:35<39:51, 95.66s/train_ba

Epoch: [11][0/26] Elapsed 1m 35s (remain 39m 51s) Loss: 0.5357 Grad: 2.5508


Train:  81%|████████████████████▎     | 21/26 [08:25<01:08, 13.80s/train_ba

Epoch: [11][20/26] Elapsed 8m 25s (remain 2m 0s) Loss: 0.6472 Grad: 3.5057
```

```
Train: 100%|████████████████████████████| 26/26 [10:11<00:00, 23.51s/train_ba


Epoch: [11][25/26] Elapsed 10m 11s (remain 0m 0s) Loss: 0.6632 Grad: 3.0698
train_epoch took 611.2627 seconds.


Validation:  11%|███                        | 1/9 [00:07<00:56,  7.10s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7622


Validation: 100%|████████████████████████████| 9/9 [03:57<00:00, 26.37s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0049
valid_epoch took 237.3576 seconds.
Epoch 11 - avg_train_loss: 0.6632  avg_val_loss: 1.0049  time: 849s


Train:   4%|█                            | 1/26 [01:30<37:43, 90.55s/train_ba

Epoch: [12][0/26] Elapsed 1m 30s (remain 37m 43s) Loss: 0.4758 Grad: 3.4338


Train:  81%|███████████████████████      | 21/26 [08:01<01:39, 19.81s/train_ba

Epoch: [12][20/26] Elapsed 8m 1s (remain 1m 54s) Loss: 0.6434 Grad: 4.1200


Train: 100%|████████████████████████████| 26/26 [10:15<00:00, 23.68s/train_ba


Epoch: [12][25/26] Elapsed 10m 15s (remain 0m 0s) Loss: 0.6233 Grad: 2.6112
train_epoch took 615.6132 seconds.


Validation:  11%|███                        | 1/9 [00:07<00:56,  7.05s/valid_ba
```

```
EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7529


Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.37s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0085
valid_epoch took 237.3705 seconds.
Epoch 12 - avg_train_loss: 0.6233  avg_val_loss: 1.0085  time: 853s


Train:   4%|▉                          | 1/26 [00:14<06:13, 14.92s/train_ba

Epoch: [13][0/26] Elapsed 0m 14s (remain 6m 13s) Loss: 0.5093 Grad: 2.9016


Train:  81%|████████████████████     | 21/26 [09:02<02:44, 32.87s/train_ba

Epoch: [13][20/26] Elapsed 9m 2s (remain 2m 9s) Loss: 0.6262 Grad: 5.2046  L


Train: 100%|█████████████████████████| 26/26 [10:16<00:00, 23.70s/train_ba


Epoch: [13][25/26] Elapsed 10m 16s (remain 0m 0s) Loss: 0.6459 Grad: 4.5728
train_epoch took 616.1525 seconds.


Validation:  11%|██                        | 1/9 [00:07<00:56,  7.11s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7639


Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.36s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0264
valid_epoch took 237.2252 seconds.
```

```
Epoch 13 - avg_train_loss: 0.6459  avg_val_loss: 1.0264  time: 853s


Train:   4%|▉                          | 1/26 [00:07<03:05,  7.44s/train_ba

Epoch: [14][0/26] Elapsed 0m 7s (remain 3m 5s) Loss: 0.6240 Grad: 2.7231  LR


Train:  81%|████████████████████     | 21/26 [09:29<02:52, 34.48s/train_ba

Epoch: [14][20/26] Elapsed 9m 29s (remain 2m 15s) Loss: 0.6627 Grad: 3.2009


Train: 100%|█████████████████████████| 26/26 [10:14<00:00, 23.63s/train_ba


Epoch: [14][25/26] Elapsed 10m 14s (remain 0m 0s) Loss: 0.6548 Grad: 5.3352
train_epoch took 614.4832 seconds.


Validation:  11%|██                        | 1/9 [00:07<00:56,  7.07s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7492


Validation: 100%|█████████████████████████| 9/9 [03:57<00:00, 26.37s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0291
valid_epoch took 237.2994 seconds.
Epoch 14 - avg_train_loss: 0.6548  avg_val_loss: 1.0291  time: 852s


Train:   4%|▉                          | 1/26 [00:08<03:26,  8.26s/train_ba

Epoch: [15][0/26] Elapsed 0m 8s (remain 3m 26s) Loss: 0.6655 Grad: 4.0483  L


Train:  81%|████████████████████     | 21/26 [08:06<01:43, 20.74s/train_ba
```

```
Epoch: [15][20/26] Elapsed 8m 6s (remain 1m 55s) Loss: 0.5894 Grad: 3.4347


Train: 100%|████████████████████████████| 26/26 [09:49<00:00, 22.69s/train_ba


Epoch: [15][25/26] Elapsed 9m 49s (remain 0m 0s) Loss: 0.6020 Grad: 2.4853
train_epoch took 589.8148 seconds.


Validation:  11%|███                        | 1/9 [00:07<00:56,  7.05s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7211


Validation: 100%|████████████████████████████| 9/9 [03:57<00:00, 26.37s/valid_ba

EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0092
valid_epoch took 237.3219 seconds.
Epoch 15 - avg_train_loss: 0.6020  avg_val_loss: 1.0092  time: 827s
Early stopping triggered at 14 epochs without improvement.




train_loop took 12654.5369 seconds.
get_result took 0.0018 seconds.
get_result took 0.0003 seconds.
========== Fold 2 Stage 1 result: 1.1740113396528333 ==========
Starting Stage 2 Training for Fold 2
Training Stage 2: Filtering data based on KL Loss < 9


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
  warnings.warn(
Train:   0%|                                    | 0/25 [00:00<?, ?train_batc
  warnings.warn(
```

```
Train:   4%|▉                              | 1/25 [00:09<03:55,  9.80s/train_ba

Epoch: [1][0/25] Elapsed 0m 9s (remain 3m 55s) Loss: 0.5566 Grad: 3.6188  LR


Train:  84%|████████████████████▉    | 21/25 [06:32<01:42, 25.59s/train_ba

Epoch: [1][20/25] Elapsed 6m 32s (remain 1m 14s) Loss: 0.6974 Grad: 3.6193


Train: 100%|█████████████████████████| 25/25 [09:41<00:00, 23.26s/train_ba


Epoch: [1][24/25] Elapsed 9m 41s (remain 0m 0s) Loss: 0.6845 Grad: 3.0806  L
train_epoch took 581.5818 seconds.


Validation:  11%|██▊                      | 1/9 [00:07<00:56,  7.12s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7661


Validation: 100%|█████████████████████████| 9/9 [03:57<00:00, 26.35s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0008
valid_epoch took 237.1482 seconds.
Epoch 1 - avg_train_loss: 0.6845  avg_val_loss: 1.0008  time: 819s


Train:   4%|▉                              | 1/25 [00:07<02:52,  7.19s/train_ba

Epoch: [2][0/25] Elapsed 0m 7s (remain 2m 52s) Loss: 0.7376 Grad: 4.7531  LR


Train:  84%|████████████████████▉    | 21/25 [07:05<02:47, 41.88s/train_ba

Epoch: [2][20/25] Elapsed 7m 5s (remain 1m 20s) Loss: 0.6091 Grad: 3.4120  L
```

```
Train: 100%|████████████████████████████| 25/25 [09:39<00:00, 23.19s/train_ba


Epoch: [2][24/25] Elapsed 9m 39s (remain 0m 0s) Loss: 0.5919 Grad: 3.5401  L
train_epoch took 579.6632 seconds.


Validation:  11%|███                        | 1/9 [00:07<00:56,  7.08s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7641


Validation: 100%|████████████████████████████| 9/9 [03:57<00:00, 26.34s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 0.9989
valid_epoch took 237.0202 seconds.
Epoch 2 - avg_train_loss: 0.5919  avg_val_loss: 0.9989  time: 817s


Train:   4%|█                            | 1/25 [00:10<04:04, 10.17s/train_ba

Epoch: [3][0/25] Elapsed 0m 10s (remain 4m 4s) Loss: 0.7327 Grad: 4.1959  LR


Train:  84%|████████████████████████     | 21/25 [07:20<01:05, 16.30s/train_ba

Epoch: [3][20/25] Elapsed 7m 20s (remain 1m 23s) Loss: 0.5733 Grad: 2.7942


Train: 100%|████████████████████████████| 25/25 [09:34<00:00, 22.99s/train_ba


Epoch: [3][24/25] Elapsed 9m 34s (remain 0m 0s) Loss: 0.5880 Grad: 3.6136  L
train_epoch took 574.8769 seconds.


Validation:  11%|███                        | 1/9 [00:07<00:56,  7.10s/valid_ba
```

```
EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7151


Validation: 100%|████████████████████████| 9/9 [03:57<00:00, 26.33s/valid_ba


EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0072
valid_epoch took 237.0130 seconds.
Epoch 3 - avg_train_loss: 0.5880  avg_val_loss: 1.0072  time: 812s


Train:   4%|▉                          | 1/25 [00:13<05:19, 13.32s/train_ba

Epoch: [4][0/25] Elapsed 0m 13s (remain 5m 19s) Loss: 0.3436 Grad: 3.1771  L


Train:  84%|████████████████████▉    | 21/25 [08:28<01:06, 16.58s/train_ba

Epoch: [4][20/25] Elapsed 8m 28s (remain 1m 36s) Loss: 0.5682 Grad: 5.9208


Train: 100%|████████████████████████| 25/25 [09:42<00:00, 23.29s/train_ba


Epoch: [4][24/25] Elapsed 9m 42s (remain 0m 0s) Loss: 0.5441 Grad: 2.4112  L
train_epoch took 582.1851 seconds.


Validation:  11%|██▋                      | 1/9 [00:07<00:56,  7.06s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7135


Validation: 100%|████████████████████████| 9/9 [03:58<00:00, 26.49s/valid_ba


EVAL: [8/9] Elapsed 3m 58s (remain 0m 0s) Loss: 1.0383
valid_epoch took 238.4150 seconds.
```

```
Epoch 4 - avg_train_loss: 0.5441  avg_val_loss: 1.0383  time: 821s


Train:   4%|▉                           | 1/25 [00:08<03:32,  8.87s/train_ba

Epoch: [5][0/25] Elapsed 0m 8s (remain 3m 32s) Loss: 0.4909 Grad: 4.4146  LR


Train:  84%|████████████████████▉     | 21/25 [08:57<01:25, 21.46s/train_ba

Epoch: [5][20/25] Elapsed 8m 57s (remain 1m 42s) Loss: 0.4666 Grad: 3.1682


Train: 100%|████████████████████████| 25/25 [09:40<00:00, 23.21s/train_ba


Epoch: [5][24/25] Elapsed 9m 40s (remain 0m 0s) Loss: 0.4659 Grad: 3.5293  L
train_epoch took 580.3399 seconds.


Validation:  11%|██▊                      | 1/9 [00:07<00:56,  7.05s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7299


Validation: 100%|████████████████████████| 9/9 [03:56<00:00, 26.33s/valid_ba


EVAL: [8/9] Elapsed 3m 56s (remain 0m 0s) Loss: 1.0200
valid_epoch took 236.9698 seconds.
Epoch 5 - avg_train_loss: 0.4659  avg_val_loss: 1.0200  time: 817s


Train:   4%|▉                           | 1/25 [00:09<03:41,  9.24s/train_ba

Epoch: [6][0/25] Elapsed 0m 9s (remain 3m 41s) Loss: 0.2256 Grad: 1.7309  LR


Train:  84%|████████████████████▉     | 21/25 [07:47<01:37, 24.49s/train_ba
```

```
Epoch: [6][20/25] Elapsed 7m 47s (remain 1m 28s) Loss: 0.4170 Grad: 4.1594


Train: 100%|████████████████████████| 25/25 [09:41<00:00, 23.26s/train_ba


Epoch: [6][24/25] Elapsed 9m 41s (remain 0m 0s) Loss: 0.4437 Grad: 3.5990  L
train_epoch took 581.5484 seconds.


Validation:  11%|██                      | 1/9 [00:07<00:56,  7.05s/valid_ba

EVAL: [0/9] Elapsed 0m 7s (remain 0m 56s) Loss: 0.7484


Validation: 100%|████████████████████| 9/9 [03:57<00:00, 26.34s/valid_ba

EVAL: [8/9] Elapsed 3m 57s (remain 0m 0s) Loss: 1.0572
valid_epoch took 237.0228 seconds.
Epoch 6 - avg_train_loss: 0.4437  avg_val_loss: 1.0572  time: 819s
Early stopping triggered at 5 epochs without improvement.




train_loop took 4904.8224 seconds.
get_result took 0.0018 seconds.
get_result took 0.0003 seconds.
========== Fold 2 Stage 2 result: 1.16407226936696 ==========
Starting Stage 1 Training for Fold 3
Training Stage 1: Using all data


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
  warnings.warn(
Train:   0%|                                  | 0/27 [00:00<?, ?train_batc
  warnings.warn(
```

```
Train:   4%|▉                              | 1/27 [00:08<03:41,  8.51s/train_ba

Epoch: [1][0/27] Elapsed 0m 8s (remain 3m 41s) Loss: 1.2258 Grad: 2.1773  LR


Train:  78%|████████████████████▉      | 21/27 [09:58<01:53, 18.93s/train_ba

Epoch: [1][20/27] Elapsed 9m 58s (remain 2m 51s) Loss: 1.3638 Grad: 2.4606


Train: 100%|███████████████████████████| 27/27 [11:25<00:00, 25.40s/train_ba


Epoch: [1][26/27] Elapsed 11m 25s (remain 0m 0s) Loss: 1.3666 Grad: 1.8512
train_epoch took 685.7310 seconds.


Validation:  11%|███                        | 1/9 [00:03<00:30,  3.83s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.4391


Validation: 100%|███████████████████████████| 9/9 [02:55<00:00, 19.45s/valid_ba


EVAL: [8/9] Elapsed 2m 55s (remain 0m 0s) Loss: 1.4768
valid_epoch took 175.0192 seconds.
Epoch 1 - avg_train_loss: 1.3666  avg_val_loss: 1.4768  time: 861s


Train:   4%|▉                              | 1/27 [00:26<11:25, 26.35s/train_ba

Epoch: [2][0/27] Elapsed 0m 26s (remain 11m 25s) Loss: 1.0654 Grad: 2.1900


Train:  78%|████████████████████▉      | 21/27 [07:48<01:59, 19.88s/train_ba

Epoch: [2][20/27] Elapsed 7m 48s (remain 2m 13s) Loss: 1.3248 Grad: 2.4736
```

```
Train: 100%|████████████████████████████| 27/27 [11:24<00:00, 25.37s/train_ba


Epoch: [2][26/27] Elapsed 11m 24s (remain 0m 0s) Loss: 1.3203 Grad: 2.3782
train_epoch took 684.8820 seconds.


Validation:  11%|███                        | 1/9 [00:03<00:30,  3.82s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.4092


Validation: 100%|████████████████████████████| 9/9 [02:54<00:00, 19.42s/valid_ba


EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.4445
valid_epoch took 174.7488 seconds.
Epoch 2 - avg_train_loss: 1.3203  avg_val_loss: 1.4445  time: 860s


Train:   4%|█                            | 1/27 [00:08<03:40,  8.47s/train_ba

Epoch: [3][0/27] Elapsed 0m 8s (remain 3m 40s) Loss: 1.1155 Grad: 1.9930  LR


Train:  78%|██████████████████████       | 21/27 [09:45<02:47, 27.99s/train_ba

Epoch: [3][20/27] Elapsed 9m 45s (remain 2m 47s) Loss: 1.2478 Grad: 2.9400


Train: 100%|████████████████████████████| 27/27 [11:40<00:00, 25.96s/train_ba


Epoch: [3][26/27] Elapsed 11m 40s (remain 0m 0s) Loss: 1.2345 Grad: 1.9056
train_epoch took 700.7963 seconds.


Validation:  11%|███                        | 1/9 [00:04<00:32,  4.09s/valid_ba
```

```
EVAL: [0/9] Elapsed 0m 4s (remain 0m 32s) Loss: 1.2761


Validation: 100%|████████████████████████| 9/9 [02:58<00:00, 19.79s/valid_ba


EVAL: [8/9] Elapsed 2m 58s (remain 0m 0s) Loss: 1.3598
valid_epoch took 178.1137 seconds.
Epoch 3 - avg_train_loss: 1.2345  avg_val_loss: 1.3598  time: 879s


Train:   4%|▉                         | 1/27 [00:08<03:50,  8.88s/train_ba

Epoch: [4][0/27] Elapsed 0m 8s (remain 3m 50s) Loss: 0.9320 Grad: 2.2016  LR


Train:  78%|███████████████████▍      | 21/27 [09:52<01:59, 19.84s/train_ba

Epoch: [4][20/27] Elapsed 9m 52s (remain 2m 49s) Loss: 1.1619 Grad: 2.9606


Train: 100%|████████████████████████| 27/27 [11:34<00:00, 25.73s/train_ba


Epoch: [4][26/27] Elapsed 11m 34s (remain 0m 0s) Loss: 1.1251 Grad: 3.2762
train_epoch took 694.8422 seconds.


Validation:  11%|██▋                      | 1/9 [00:03<00:31,  3.90s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 31s) Loss: 1.2105


Validation: 100%|████████████████████████| 9/9 [02:55<00:00, 19.50s/valid_ba


EVAL: [8/9] Elapsed 2m 55s (remain 0m 0s) Loss: 1.2832
valid_epoch took 175.5358 seconds.
```

```
Epoch 4 - avg_train_loss: 1.1251  avg_val_loss: 1.2832  time: 870s


Train:   4%|▉                         | 1/27 [02:27<1:03:56, 147.56s/train_ba

Epoch: [5][0/27] Elapsed 2m 27s (remain 63m 56s) Loss: 0.9635 Grad: 2.7679


Train:  78%|████████████████████      | 21/27 [08:23<03:52, 38.68s/train_ba

Epoch: [5][20/27] Elapsed 8m 23s (remain 2m 23s) Loss: 1.0271 Grad: 2.3466


Train: 100%|██████████████████████████| 27/27 [11:29<00:00, 25.53s/train_ba


Epoch: [5][26/27] Elapsed 11m 29s (remain 0m 0s) Loss: 1.0031 Grad: 3.1115
train_epoch took 689.4316 seconds.


Validation:  11%|██                       | 1/9 [00:03<00:30,  3.87s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.2069


Validation: 100%|██████████████████████████| 9/9 [02:54<00:00, 19.41s/valid_ba


EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.2388
valid_epoch took 174.7335 seconds.
Epoch 5 - avg_train_loss: 1.0031  avg_val_loss: 1.2388  time: 864s


Train:   4%|▉                           | 1/27 [00:25<11:13, 25.89s/train_ba

Epoch: [6][0/27] Elapsed 0m 25s (remain 11m 13s) Loss: 0.9775 Grad: 2.6725


Train:  78%|████████████████████      | 21/27 [08:44<03:31, 35.19s/train_ba
```

```
Epoch: [6][20/27] Elapsed 8m 44s (remain 2m 29s) Loss: 0.9098 Grad: 3.9440


Train: 100%|██████████████████████████| 27/27 [11:26<00:00, 25.43s/train_ba


Epoch: [6][26/27] Elapsed 11m 26s (remain 0m 0s) Loss: 0.9046 Grad: 3.2372
train_epoch took 686.6481 seconds.


Validation:  11%|███                       | 1/9 [00:03<00:30,  3.77s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.2115


Validation: 100%|████████████████████████| 9/9 [02:54<00:00, 19.41s/valid_ba


EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.1607
valid_epoch took 174.7055 seconds.
Epoch 6 - avg_train_loss: 0.9046  avg_val_loss: 1.1607  time: 861s


Train:   4%|█                           | 1/27 [00:25<11:06, 25.62s/train_ba

Epoch: [7][0/27] Elapsed 0m 25s (remain 11m 6s) Loss: 0.7485 Grad: 2.3626  L


Train:  78%|████████████████████      | 21/27 [09:58<02:56, 29.45s/train_ba

Epoch: [7][20/27] Elapsed 9m 58s (remain 2m 51s) Loss: 0.7615 Grad: 6.5665


Train: 100%|██████████████████████████| 27/27 [11:27<00:00, 25.45s/train_ba


Epoch: [7][26/27] Elapsed 11m 27s (remain 0m 0s) Loss: 0.7736 Grad: 3.9524
train_epoch took 687.1457 seconds.
```

```
Validation:  11%|██                    | 1/9 [00:03<00:30,  3.85s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.3231

Validation: 100%|████████████████████| 9/9 [02:54<00:00, 19.43s/valid_ba

EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.1719
valid_epoch took 174.8503 seconds.
Epoch 7 - avg_train_loss: 0.7736  avg_val_loss: 1.1719  time: 862s

Train:   4%|▊                      | 1/27 [00:10<04:38, 10.69s/train_ba

Epoch: [8][0/27] Elapsed 0m 10s (remain 4m 38s) Loss: 0.8291 Grad: 2.8369  L

Train:  78%|███████████████▌     | 21/27 [09:54<02:16, 22.81s/train_ba

Epoch: [8][20/27] Elapsed 9m 54s (remain 2m 49s) Loss: 0.7443 Grad: 3.4980

Train: 100%|████████████████████| 27/27 [11:28<00:00, 25.49s/train_ba

Epoch: [8][26/27] Elapsed 11m 28s (remain 0m 0s) Loss: 0.7481 Grad: 2.0705
train_epoch took 688.3310 seconds.

Validation:  11%|██                    | 1/9 [00:03<00:31,  3.88s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 31s) Loss: 1.2241

Validation: 100%|████████████████████| 9/9 [02:55<00:00, 19.47s/valid_ba
```

```
EVAL: [8/9] Elapsed 2m 55s (remain 0m 0s) Loss: 1.1618
valid_epoch took 175.2389 seconds.
Epoch 8 - avg_train_loss: 0.7481  avg_val_loss: 1.1618  time: 864s


Train:   4%|▉                          | 1/27 [00:08<03:41,  8.54s/train_ba

Epoch: [9][0/27] Elapsed 0m 8s (remain 3m 41s) Loss: 0.6289 Grad: 2.7352  LR


Train:  78%|███████████████████      | 21/27 [10:02<02:35, 25.84s/train_ba

Epoch: [9][20/27] Elapsed 10m 2s (remain 2m 52s) Loss: 0.7287 Grad: 2.6575


Train: 100%|████████████████████████| 27/27 [11:26<00:00, 25.42s/train_ba


Epoch: [9][26/27] Elapsed 11m 26s (remain 0m 0s) Loss: 0.7248 Grad: 3.3739
train_epoch took 686.2162 seconds.


Validation:  11%|██                     | 1/9 [00:03<00:31,  3.92s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 31s) Loss: 1.1849


Validation: 100%|████████████████████| 9/9 [02:54<00:00, 19.42s/valid_ba


EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.1400
valid_epoch took 174.7570 seconds.
Epoch 9 - avg_train_loss: 0.7248  avg_val_loss: 1.1400  time: 861s


Train:   4%|▉                          | 1/27 [01:30<39:14, 90.57s/train_ba

Epoch: [10][0/27] Elapsed 1m 30s (remain 39m 14s) Loss: 0.6863 Grad: 3.7450
```

```
Train:  78%|███████████████████      | 21/27 [07:06<01:08, 11.43s/train_ba

Epoch: [10][20/27] Elapsed 7m 6s (remain 2m 1s) Loss: 0.6336 Grad: 3.2203  L

Train: 100%|████████████████████████| 27/27 [11:26<00:00, 25.42s/train_ba

Epoch: [10][26/27] Elapsed 11m 26s (remain 0m 0s) Loss: 0.6551 Grad: 3.3054
train_epoch took 686.3678 seconds.

Validation:  11%|██                       | 1/9 [00:03<00:31,  3.89s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 31s) Loss: 1.1241

Validation: 100%|████████████████████████| 9/9 [02:54<00:00, 19.43s/valid_ba

EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.1494
valid_epoch took 174.8686 seconds.
Epoch 10 - avg_train_loss: 0.6551  avg_val_loss: 1.1494  time: 861s

Train:   4%|▉                          | 1/27 [00:11<04:56, 11.39s/train_ba

Epoch: [11][0/27] Elapsed 0m 11s (remain 4m 56s) Loss: 0.7954 Grad: 4.7780

Train:  78%|███████████████████      | 21/27 [10:20<02:03, 20.57s/train_ba

Epoch: [11][20/27] Elapsed 10m 20s (remain 2m 57s) Loss: 0.6514 Grad: 3.5566

Train: 100%|████████████████████████| 27/27 [11:24<00:00, 25.36s/train_ba
```

```
Epoch: [11][26/27] Elapsed 11m 24s (remain 0m 0s) Loss: 0.6545 Grad: 4.3495
train_epoch took 684.6309 seconds.


Validation:  11%|██                    | 1/9 [00:03<00:30,  3.84s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.2198


Validation: 100%|████████████████████| 9/9 [02:54<00:00, 19.41s/valid_ba

EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.1502
valid_epoch took 174.7012 seconds.
Epoch 11 - avg_train_loss: 0.6545  avg_val_loss: 1.1502  time: 859s
Early stopping triggered at 10 epochs without improvement.




train_loop took 9503.5504 seconds.
get_result took 0.0018 seconds.
get_result took 0.0003 seconds.
========== Fold 3 Stage 1 result: 1.3382047760292197 ==========
Starting Stage 2 Training for Fold 3
Training Stage 2: Filtering data based on KL Loss < 9


/opt/anaconda3/envs/myenv/lib/python3.11/site-packages/torch/amp/grad_scaler
  warnings.warn(
Train:   0%|                                | 0/25 [00:00<?, ?train_batc
  warnings.warn(
Train:   4%|█                           | 1/25 [00:11<04:47, 11.96s/train_ba

Epoch: [1][0/25] Elapsed 0m 11s (remain 4m 47s) Loss: 0.5816 Grad: 2.5870  L


Train:  84%|██████████████████    | 21/25 [10:26<03:08, 47.24s/train_ba
```

```
Epoch: [1][20/25] Elapsed 10m 26s (remain 1m 59s) Loss: 0.6724 Grad: 4.2521

Train: 100%|████████████████████████| 25/25 [11:10<00:00, 26.83s/train_ba

Epoch: [1][24/25] Elapsed 11m 10s (remain 0m 0s) Loss: 0.6611 Grad: 3.8444
train_epoch took 670.7317 seconds.

Validation:  11%|██                      | 1/9 [00:03<00:30,  3.85s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 1.1106

Validation: 100%|████████████████████████| 9/9 [02:54<00:00, 19.44s/valid_ba

EVAL: [8/9] Elapsed 2m 54s (remain 0m 0s) Loss: 1.1560
valid_epoch took 174.9272 seconds.
Epoch 1 - avg_train_loss: 0.6611  avg_val_loss: 1.1560  time: 846s

Train:   4%|▉                         | 1/25 [00:13<05:19, 13.29s/train_ba

Epoch: [2][0/25] Elapsed 0m 13s (remain 5m 19s) Loss: 0.4434 Grad: 1.9202  L

Train:  84%|████████████████████    | 21/25 [10:24<02:02, 30.68s/train_ba

Epoch: [2][20/25] Elapsed 10m 24s (remain 1m 59s) Loss: 0.5994 Grad: 2.9423

Train: 100%|████████████████████████| 25/25 [11:19<00:00, 27.18s/train_ba

Epoch: [2][24/25] Elapsed 11m 19s (remain 0m 0s) Loss: 0.5996 Grad: 2.6553
train_epoch took 679.4581 seconds.
```

```
Validation:  11%|██                    | 1/9 [00:03<00:31,  3.97s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 31s) Loss: 1.0424

Validation: 100%|████████████████████| 9/9 [02:55<00:00, 19.54s/valid_ba

EVAL: [8/9] Elapsed 2m 55s (remain 0m 0s) Loss: 1.1636
valid_epoch took 175.8371 seconds.
Epoch 2 - avg_train_loss: 0.5996  avg_val_loss: 1.1636  time: 855s

Train:   4%|▊                      | 1/25 [00:07<03:11,  8.00s/train_ba

Epoch: [3][0/25] Elapsed 0m 7s (remain 3m 11s) Loss: 0.3621 Grad: 3.1818  LR

Train:  84%|████████████████▊    | 21/25 [08:03<01:10, 17.63s/train_ba

Epoch: [3][20/25] Elapsed 8m 3s (remain 1m 32s) Loss: 0.5802 Grad: 4.0230  L

Train: 100%|████████████████████| 25/25 [11:00<00:00, 26.42s/train_ba

Epoch: [3][24/25] Elapsed 11m 0s (remain 0m 0s) Loss: 0.6158 Grad: 4.7840  L
train_epoch took 660.5254 seconds.

Validation:  11%|██                    | 1/9 [00:03<00:31,  3.94s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 31s) Loss: 1.2980

Validation: 100%|████████████████████| 9/9 [02:55<00:00, 19.52s/valid_ba
```

```
EVAL: [8/9] Elapsed 2m 55s (remain 0m 0s) Loss: 1.1251
valid_epoch took 175.7075 seconds.
Epoch 3 - avg_train_loss: 0.6158  avg_val_loss: 1.1251  time: 836s


Train:   4%|▉                          | 1/25 [00:08<03:16,  8.17s/train_ba

Epoch: [4][0/25] Elapsed 0m 8s (remain 3m 16s) Loss: 0.6261 Grad: 3.2917  LR


Train:  84%|████████████████████     | 21/25 [10:09<01:26, 21.72s/train_ba

Epoch: [4][20/25] Elapsed 10m 9s (remain 1m 56s) Loss: 0.4593 Grad: 2.2523


Train: 100%|████████████████████████| 25/25 [11:10<00:00, 26.83s/train_ba


Epoch: [4][24/25] Elapsed 11m 10s (remain 0m 0s) Loss: 0.4789 Grad: 3.6552
train_epoch took 670.7896 seconds.


Validation:  11%|██                      | 1/9 [00:03<00:30,  3.84s/valid_ba

EVAL: [0/9] Elapsed 0m 3s (remain 0m 30s) Loss: 0.9722


Validation: 100%|████████████████████| 9/9 [02:55<00:00, 19.49s/valid_ba


EVAL: [8/9] Elapsed 2m 55s (remain 0m 0s) Loss: 1.1881
valid_epoch took 175.4069 seconds.
Epoch 4 - avg_train_loss: 0.4789  avg_val_loss: 1.1881  time: 846s


Train:   4%|▉                          | 1/25 [00:08<03:33,  8.91s/train_ba

Epoch: [5][0/25] Elapsed 0m 8s (remain 3m 33s) Loss: 0.5843 Grad: 4.9719  LR
```

```
Train:  84%|████████████████████▉    | 21/25 [07:56<01:50, 27.67s/train_ba

Epoch: [5][20/25] Elapsed 7m 56s (remain 1m 30s) Loss: 0.5069 Grad: 5.0968


Train: 100%|█████████████████████████| 25/25 [11:34<00:00, 27.79s/train_ba

Epoch: [5][24/25] Elapsed 11m 34s (remain 0m 0s) Loss: 0.4844 Grad: 2.2097
train_epoch took 694.8448 seconds.


Validation:  11%|██                    | 1/9 [00:04<00:33,  4.19s/valid_ba

EVAL: [0/9] Elapsed 0m 4s (remain 0m 33s) Loss: 1.1410


Validation: 100%|█████████████████████| 9/9 [02:59<00:00, 19.90s/valid_ba


EVAL: [8/9] Elapsed 2m 59s (remain 0m 0s) Loss: 1.1689
valid_epoch took 179.1185 seconds.
Epoch 5 - avg_train_loss: 0.4844  avg_val_loss: 1.1689  time: 874s


Train:   4%|▉                         | 1/25 [00:15<06:20, 15.85s/train_ba

Epoch: [6][0/25] Elapsed 0m 15s (remain 6m 20s) Loss: 0.4115 Grad: 2.3254  L


Train:  84%|████████████████████▉    | 21/25 [10:21<02:55, 43.89s/train_ba

Epoch: [6][20/25] Elapsed 10m 21s (remain 1m 58s) Loss: 0.4383 Grad: 4.1395


Train: 100%|█████████████████████████| 25/25 [11:28<00:00, 27.55s/train_ba
```

```
Epoch: [6][24/25] Elapsed 11m 28s (remain 0m 0s) Loss: 0.4259 Grad: 3.7884
train_epoch took 688.6849 seconds.


Validation:  11%|██                    | 1/9 [00:04<00:33,  4.21s/valid_ba

EVAL: [0/9] Elapsed 0m 4s (remain 0m 33s) Loss: 1.4032


Validation: 100%|████████████████████| 9/9 [02:59<00:00, 19.91s/valid_ba

EVAL: [8/9] Elapsed 2m 59s (remain 0m 0s) Loss: 1.2365
valid_epoch took 179.2121 seconds.
Epoch 6 - avg_train_loss: 0.4259  avg_val_loss: 1.2365  time: 868s
Early stopping triggered at 5 epochs without improvement.




train_loop took 5126.3617 seconds.
get_result took 0.0016 seconds.
get_result took 0.0003 seconds.
========== Fold 3 Stage 2 result: 1.3233752799339686 ==========
get_result took 0.0005 seconds.
========== CV: 1.2527229396406405 ==========
```