model 1

---

imports

```
import tensorflow as tf
from keras.models import load_model
from keras.callbacks import ModelCheckpoint, EarlyStopping
from keras_tqdm import TQDMNotebookCallback
import numpy as np
from keras_tqdm import TQDMNotebookCallback
import nltk
import xml.etree.ElementTree as ET
import pandas as pd
import os
import string
from nltk.tokenize import TreebankWordTokenizer
from numpy.random import random_sample
import re
import pickle
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report

from keras.layers import Embedding, Flatten,LSTM
from keras.layers.convolutional import Conv2D, MaxPooling2D
from keras.utils import to_categorical
from keras.models import Sequential, Model
from keras.layers import Dense, Dropout, Activation,  Input, merge,Conv1D,MaxPooling1D,GlobalMaxPooling1D,Convolution1D
from keras import regularizers
from sklearn.metrics import precision_recall_fscore_support
from sklearn.model_selection import StratifiedKFold
import matplotlib.pyplot as plt
from keras.layers import Concatenate, concatenate
from keras import backend as K
from keras.layers import multiply
from keras.layers import merge
from keras.layers.core import *
from keras.layers.recurrent import LSTM
from keras.models import *
from keras.regularizers import l2
random_seed=1337
```

```
Using TensorFlow backend.
```

Define Callback functions to generate Measures

```
from keras import backend as K

def f1(y_true, y_pred):
    def recall(y_true, y_pred):
        true_positives = K.sum(K.round(K.clip(y_true * y_pred, 0, 1)))
        possible_positives = K.sum(K.round(K.clip(y_true, 0, 1)))
        recall = true_positives / (possible_positives + K.epsilon())
        return recall

    def precision(y_true, y_pred):
        true_positives = K.sum(K.round(K.clip(y_true * y_pred, 0, 1)))
        predicted_positives = K.sum(K.round(K.clip(y_pred, 0, 1)))
        precision = true_positives / (predicted_positives + K.epsilon())
        return precision
    precision = precision(y_true, y_pred)
    recall = recall(y_true, y_pred)
    return 2*((precision*recall)/(precision+recall+K.epsilon()))
```

Experiments to reproduce the results of Table 9

Load pre procssed Data

```
with open('../data/pickles/train_and_test_data_sentences_snp_2class.pickle', 'rb') as handle:
#with open('../../SNP-Disease/train_and_test_data_sentences_snp_2classWiki.pickle', 'rb') as handle:

    W_train = pickle.load(handle)
    d1_train = pickle.load(handle)
    d2_train = pickle.load(handle)
    Y_train = pickle.load(handle)
    Tr_word_list = pickle.load(handle)

    W_test = pickle.load(handle)
    d1_test = pickle.load(handle)
    d2_test = pickle.load(handle)
    Y_test = pickle.load(handle)
    Te_word_list = pickle.load(handle)


    word_vectors = pickle.load(handle)
    word_dict = pickle.load(handle)
    d1_dict = pickle.load(handle)
    d2_dict = pickle.load(handle)
    label_dict = pickle.load(handle)
    MAX_SEQUENCE_LENGTH = pickle.load(handle)
```

Prepare Word Embedding Layer

```
EMBEDDING_DIM=word_vectors.shape[1]
embedding_matrix=word_vectors

def create_embedding_layer(l2_reg=0.1,use_pretrained=True,is_trainable=False):

    if use_pretrained:
        return Embedding(len(word_dict) ,EMBEDDING_DIM,weights=[embedding_matrix],input_length=MAX_SEQUENCE_LENGTH,trainable=is_trainable,em
    else:
        return Embedding(len(word_dict) ,EMBEDDING_DIM,input_length=MAX_SEQUENCE_LENGTH)



INPUT_DIM = 2
TIME_STEPS = MAX_SEQUENCE_LENGTH


def attentionNew(inputs):
    inputs = Lambda(lambda x: tf.keras.backend.sigmoid(x))(inputs)
    input_dim = int(inputs.shape[2])
    a = Permute((2, 1))(inputs)
    a = Dense(TIME_STEPS, activation='softmax')(a)
    a_probs = Permute((2, 1))(a)
    output_attention_mul = multiply([inputs, a_probs])
    output_attention_mul = Lambda(lambda x: tf.keras.backend.sigmoid(x))(output_attention_mul)
    return output_attention_mul
```

Create the Model

```
from keras.optimizers import Adam
def build_model():
    sequence_input = Input(shape=(MAX_SEQUENCE_LENGTH,), dtype='int32')
    embedding_layer = create_embedding_layer(use_pretrained=True, is_trainable=False)
    embedded_sequences = embedding_layer(sequence_input)

    # First Conv1D Layer
    x = Conv1D(256, 7, activation='relu')(embedded_sequences)
```

```
    x = MaxPooling1D(3)(x)
    x = Dropout(0.4)(x)

    # Second Conv1D Layer
    x = Conv1D(128, 5, activation='relu')(x)
    x = MaxPooling1D(3)(x)
    x = Dropout(0.4)(x)

    conv_sequence = GlobalMaxPooling1D()(x)  # GlobalMaxPooling after all CNN layers

    # Bidirectional RNN layers
    forward = LSTM(100, return_sequences=True, recurrent_dropout=0.05)(embedded_sequences)
    backward = LSTM(100, return_sequences=True, go_backwards=True, recurrent_dropout=0.05)(embedded_sequences)
    lstm_gru_sequence = concatenate([forward, backward], axis=-1)

    # Apply attention mechanism
    attention_output = attentionNew(lstm_gru_sequence)
    attention_pooled = GlobalMaxPooling1D()(attention_output)

    # Merge CNN and attention-enhanced RNN outputs
    merge = concatenate([conv_sequence, attention_pooled])

    # Fully connected layers
    x = Dense(256, activation='relu', kernel_regularizer=regularizers.l2(0.05))(merge)
    x = Dropout(0.4)(x)
    preds = Dense(2, activation='softmax')(x)

    # Compile model
    optimizer = Adam(learning_rate=0.001)
    model = Model(sequence_input, preds)
    model.compile(loss='binary_crossentropy', optimizer=optimizer, metrics=['acc', f1])

    return model



model = build_model()
model.summary()
```

```
Model: "model_1"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_1 (InputLayer)            (None, 91)           0
__________________________________________________________________________________________________
embedding_1 (Embedding)         (None, 91, 200)      555000      input_1[0][0]
__________________________________________________________________________________________________
lstm_1 (LSTM)                   (None, 91, 100)      120400      embedding_1[0][0]
__________________________________________________________________________________________________
lstm_2 (LSTM)                   (None, 91, 100)      120400      embedding_1[0][0]
__________________________________________________________________________________________________
concatenate_1 (Concatenate)     (None, 91, 200)      0           lstm_1[0][0]
                                                                 lstm_2[0][0]
__________________________________________________________________________________________________
conv1d_1 (Conv1D)               (None, 85, 256)      358656      embedding_1[0][0]
__________________________________________________________________________________________________
lambda_1 (Lambda)               (None, 91, 200)      0           concatenate_1[0][0]
__________________________________________________________________________________________________
max_pooling1d_1 (MaxPooling1D)  (None, 28, 256)      0           conv1d_1[0][0]
__________________________________________________________________________________________________
permute_1 (Permute)             (None, 200, 91)      0           lambda_1[0][0]
__________________________________________________________________________________________________
dropout_1 (Dropout)             (None, 28, 256)      0           max_pooling1d_1[0][0]
__________________________________________________________________________________________________
dense_1 (Dense)                 (None, 200, 91)      8372        permute_1[0][0]
__________________________________________________________________________________________________
conv1d_2 (Conv1D)               (None, 24, 128)      163968      dropout_1[0][0]
__________________________________________________________________________________________________
permute_2 (Permute)             (None, 91, 200)      0           dense_1[0][0]
__________________________________________________________________________________________________
max_pooling1d_2 (MaxPooling1D)  (None, 8, 128)       0           conv1d_2[0][0]
__________________________________________________________________________________________________
multiply_1 (Multiply)           (None, 91, 200)      0           lambda_1[0][0]
                                                                 permute_2[0][0]
__________________________________________________________________________________________________
dropout_2 (Dropout)             (None, 8, 128)       0           max_pooling1d_2[0][0]
__________________________________________________________________________________________________
lambda_2 (Lambda)               (None, 91, 200)      0           multiply_1[0][0]
__________________________________________________________________________________________________
global_max_pooling1d_1 (GlobalM (None, 128)          0           dropout_2[0][0]
__________________________________________________________________________________________________
```

```
global_max_pooling1d_2 (GlobalM (None, 200)         0           lambda_2[0][0]
__________________________________________________________________________________________________
concatenate_2 (Concatenate)     (None, 328)         0           global_max_pooling1d_1[0][0]
                                                                 global_max_pooling1d_2[0][0]
__________________________________________________________________________________________________
dense_2 (Dense)                 (None, 256)         84224       concatenate_2[0][0]
__________________________________________________________________________________________________
dropout_3 (Dropout)             (None, 256)         0           dense_2[0][0]
__________________________________________________________________________________________________
dense_3 (Dense)                 (None, 2)           514         dropout_3[0][0]
==================================================================================================
Total params: 1,411,534
Trainable params: 856,534
Non-trainable params: 555,000
__________________________________________________________________________________________________
```

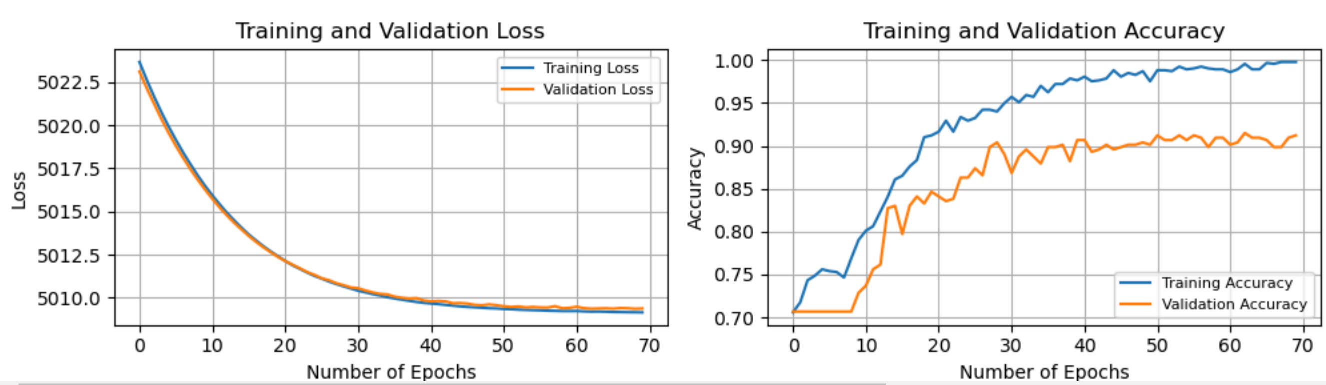
Run the Evaluation on the test dataset

```
param='macro'
epochs =70
batch_size = 32
history=model.fit(W_train, Y_train,epochs=epochs,validation_data=(W_test,Y_test), batch_size=batch_size,verbose=1)
```

```
Train on 935 samples, validate on 365 samples
Epoch 1/70
935/935 [==============================] - 2s 3ms/step - loss: 5023.6647 - acc: 0.7059 - f1: 0.6949 - val_loss: 5023.1115 - val_acc:
Epoch 2/70
935/935 [==============================] - 3s 3ms/step - loss: 5022.6282 - acc: 0.7176 - f1: 0.7213 - val_loss: 5022.1397 - val_acc:
Epoch 3/70
935/935 [==============================] - 3s 3ms/step - loss: 5021.6634 - acc: 0.7433 - f1: 0.7500 - val_loss: 5021.2463 - val_acc:
Epoch 4/70
935/935 [==============================] - 3s 3ms/step - loss: 5020.7641 - acc: 0.7487 - f1: 0.7515 - val_loss: 5020.3706 - val_acc:
Epoch 5/70
935/935 [==============================] - 3s 3ms/step - loss: 5019.9258 - acc: 0.7561 - f1: 0.7588 - val_loss: 5019.5534 - val_acc:
Epoch 6/70
935/935 [==============================] - 3s 3ms/step - loss: 5019.1490 - acc: 0.7540 - f1: 0.7455 - val_loss: 5018.8039 - val_acc:
Epoch 7/70
935/935 [==============================] - 3s 3ms/step - loss: 5018.4159 - acc: 0.7529 - f1: 0.7482 - val_loss: 5018.1015 - val_acc:
Epoch 8/70
935/935 [==============================] - 3s 3ms/step - loss: 5017.7388 - acc: 0.7465 - f1: 0.7457 - val_loss: 5017.4493 - val_acc:
Epoch 9/70
935/935 [==============================] - 3s 3ms/step - loss: 5017.0862 - acc: 0.7690 - f1: 0.7638 - val_loss: 5016.8374 - val_acc:
Epoch 10/70
935/935 [==============================] - 3s 3ms/step - loss: 5016.4904 - acc: 0.7904 - f1: 0.7921 - val_loss: 5016.2697 - val_acc:
Epoch 11/70
935/935 [==============================] - 3s 3ms/step - loss: 5015.9368 - acc: 0.8011 - f1: 0.8025 - val_loss: 5015.7382 - val_acc:
Epoch 12/70
935/935 [==============================] - 3s 3ms/step - loss: 5015.4262 - acc: 0.8064 - f1: 0.8115 - val_loss: 5015.2357 - val_acc:
Epoch 13/70
935/935 [==============================] - 3s 3ms/step - loss: 5014.9288 - acc: 0.8235 - f1: 0.8244 - val_loss: 5014.7739 - val_acc:
Epoch 14/70
935/935 [==============================] - 3s 3ms/step - loss: 5014.4881 - acc: 0.8406 - f1: 0.8411 - val_loss: 5014.3467 - val_acc:
Epoch 15/70
935/935 [==============================] - 3s 3ms/step - loss: 5014.0656 - acc: 0.8610 - f1: 0.8609 - val_loss: 5013.9601 - val_acc:
Epoch 16/70
935/935 [==============================] - 3s 3ms/step - loss: 5013.6705 - acc: 0.8652 - f1: 0.8613 - val_loss: 5013.5864 - val_acc:
Epoch 17/70
935/935 [==============================] - 3s 3ms/step - loss: 5013.3233 - acc: 0.8759 - f1: 0.8717 - val_loss: 5013.2443 - val_acc:
Epoch 18/70
935/935 [==============================] - 3s 3ms/step - loss: 5012.9941 - acc: 0.8834 - f1: 0.8827 - val_loss: 5012.9284 - val_acc:
Epoch 19/70
935/935 [==============================] - 3s 3ms/step - loss: 5012.6926 - acc: 0.9102 - f1: 0.9088 - val_loss: 5012.6458 - val_acc:
Epoch 20/70
935/935 [==============================] - 3s 3ms/step - loss: 5012.4137 - acc: 0.9123 - f1: 0.9146 - val_loss: 5012.3743 - val_acc:
Epoch 21/70
935/935 [==============================] - 3s 3ms/step - loss: 5012.1405 - acc: 0.9166 - f1: 0.9187 - val_loss: 5012.1335 - val_acc:
Epoch 22/70
935/935 [==============================] - 3s 3ms/step - loss: 5011.8899 - acc: 0.9294 - f1: 0.9312 - val_loss: 5011.9222 - val_acc:
Epoch 23/70
935/935 [==============================] - 3s 3ms/step - loss: 5011.6857 - acc: 0.9166 - f1: 0.9187 - val_loss: 5011.7147 - val_acc:
Epoch 24/70
935/935 [==============================] - 3s 3ms/step - loss: 5011.4714 - acc: 0.9337 - f1: 0.9317 - val_loss: 5011.4957 - val_acc:
Epoch 25/70
935/935 [==============================] - 3s 3ms/step - loss: 5011.2751 - acc: 0.9294 - f1: 0.9312 - val_loss: 5011.3360 - val_acc:
Epoch 26/70
935/935 [==============================] - 3s 4ms/step - loss: 5011.1183 - acc: 0.9326 - f1: 0.9344 - val_loss: 5011.1361 - val_acc:
Epoch 27/70
935/935 [==============================] - 3s 3ms/step - loss: 5010.9499 - acc: 0.9422 - f1: 0.9438 - val_loss: 5011.0256 - val_acc:
Epoch 28/70
```

```
935/935 [==============================] - 3s 3ms/step - loss: 5010.8037 - acc: 0.9422 - f1: 0.9438 - val_loss: 5010.8508 - val_acc:
```

```
import matplotlib.pyplot as plt

# Training & Validation accuracy
train_loss = history.history['loss']
val_loss = history.history['val_loss']
train_acc = history.history['acc']
val_acc = history.history['val_acc']
epochs = len(train_loss)

xc = range(epochs)

plt.figure(figsize=(10, 3))

# Loss subplot
plt.subplot(1, 2, 1)
plt.plot(xc, train_loss, label='Training Loss')
plt.plot(xc, val_loss, label='Validation Loss')
plt.xlabel('Number of Epochs', fontsize=10)
plt.ylabel('Loss', fontsize=10)
plt.title('Training and Validation Loss', fontsize=12)
plt.legend(fontsize=8)
plt.grid(True)

# Accuracy subplot
plt.subplot(1, 2, 2)
plt.plot(xc, train_acc, label='Training Accuracy')
plt.plot(xc, val_acc, label='Validation Accuracy')
plt.xlabel('Number of Epochs', fontsize=10)
plt.ylabel('Accuracy', fontsize=10)
plt.title('Training and Validation Accuracy', fontsize=12)
plt.legend(fontsize=8, loc='lower right')  # Change position to lower right
plt.grid(True)

plt.tight_layout()
plt.show()
```

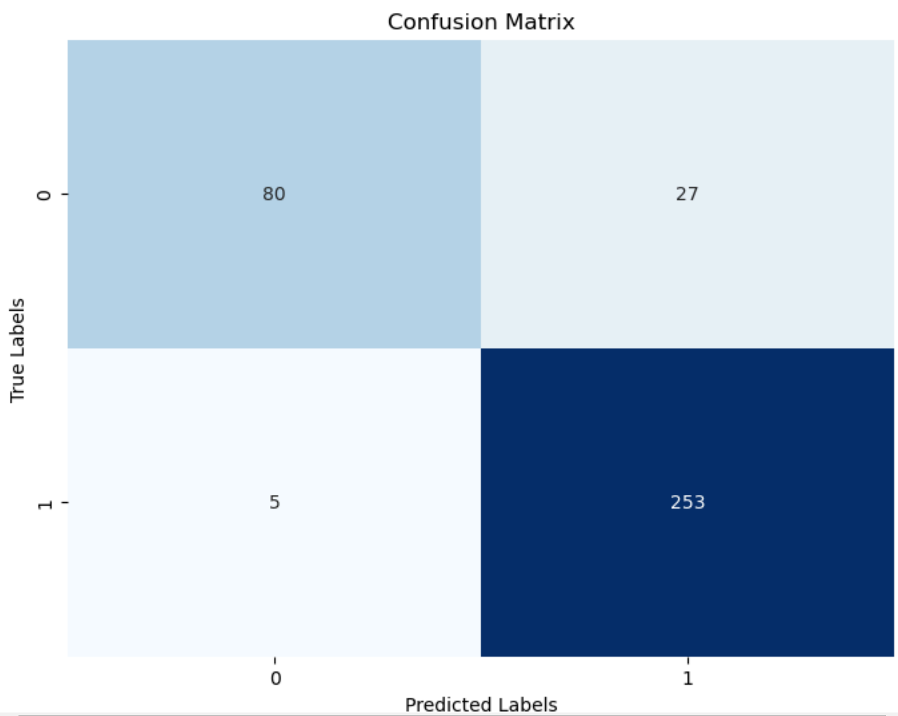



```
predicted = np.argmax(model.predict(W_test), axis=1)
y_test_to_label= np.argmax(Y_test, axis=1)
prec, reca, fscore, sup = precision_recall_fscore_support(y_test_to_label, predicted, average=param)


# Generate the classification report as a dictionary
report_dict = classification_report(y_test_to_label, predicted, output_dict=True)

# Create a new dictionary to hold the formatted values
formatted_report_dict = {}

# Iterate over the items in the report dictionary
for key, value in report_dict.items():
    if isinstance(value, dict):
        # Format the nested dictionary values
        formatted_report_dict[key] = {sub_key: f"{sub_value:.4f}" for sub_key, sub_value in value.items()}
    else:
        # Format the top-level dictionary values
        formatted_report_dict[key] = f"{value:.4f}"
```

```
# Create a string representation of the formatted dictionary
formatted_report_str = classification_report(y_test_to_label, predicted, digits=4)

# Print the formatted classification report
print(formatted_report_str)


print(" Precision:{:.2f}% Recall:{:.2f}% Fscore:{:.2f}% ".format(prec*100, reca*100, fscore*100))
```

```
              precision    recall  f1-score   support

           0     0.9412    0.7477    0.8333       107
           1     0.9036    0.9806    0.9405       258

    accuracy                         0.9123       365
   macro avg     0.9224    0.8641    0.8869       365
weighted avg     0.9146    0.9123    0.9091       365

 Precision:92.24% Recall:86.41% Fscore:88.69%
```

```
from sklearn.metrics import confusion_matrix
import seaborn as sns

# Get the predicted labels
predicted_labels = np.argmax(model.predict(W_test), axis=1)

# Create the confusion matrix
cm = confusion_matrix(np.argmax(Y_test, axis=1), predicted_labels)

# Plot the confusion matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', cbar=False)
plt.xlabel('Predicted Labels')
plt.ylabel('True Labels')
plt.title('Confusion Matrix')
plt.show()
```



Start coding or generate with AI.