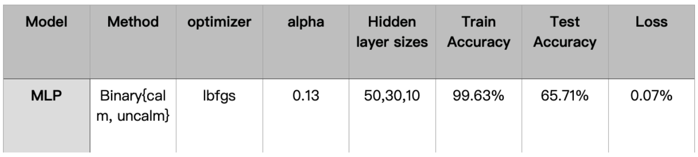
Experiment Results

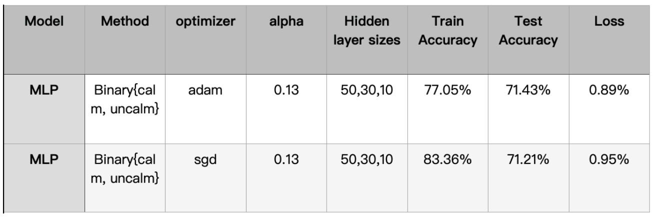
Our experiment starts with an MLP model, a simple approach to train and test the dataset, and we use a binary approach to separate emotions from the dataset into calm and uncalm. The results in Table 1 show the train accuracy and test accuracy for the lbfgs optimizer.



Model	Method	optimizer	alpha	Hidden layer sizes	Train Accuracy	Test Accuracy	Loss
MLP	Binary{calm, uncalm}	lbfgs	0.13	50,30,10	99.63%	65.71%	0.07%

Table 1

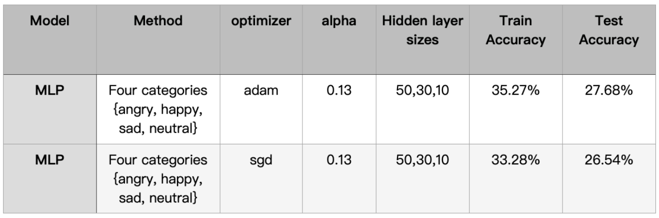
Based on results in table 1, our MLP model did not perform well for test accuracy because our model was over-fitting, therefore, we changed the optimizer in the MLP model, and the results in Table 2 show the performance of train accuracy and test accuracy in MLP for different optimizers.



Model	Method	optimizer	alpha	Hidden layer sizes	Train Accuracy	Test Accuracy	Loss
MLP	Binary{calm, uncalm}	adam	0.13	50,30,10	77.05%	71.43%	0.89%
MLP	Binary{calm, uncalm}	sgd	0.13	50,30,10	83.36%	71.21%	0.95%

Table 2

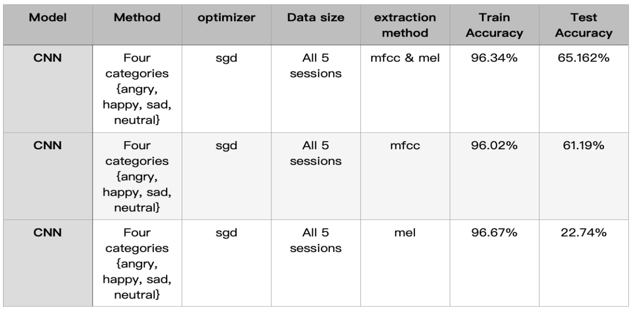
Comparing the results in Table 1 and Table 2, test accuracy is increased because adam and sgd optimizer is suitable for large dataset. Then, we use four categories to separate emotions from dataset into angry, happy, sad, and neutral, and the results in Table 3 show that our train accuracy and test accuracy for 4-categories method in MLP Model.



Model	Method	optimizer	alpha	Hidden layer sizes	Train Accuracy	Test Accuracy
MLP	Four categories {angry, happy, sad, neutral}	adam	0.13	50,30,10	35.27%	27.68%
MLP	Four categories {angry, happy, sad, neutral}	sgd	0.13	50,30,10	33.28%	26.54%

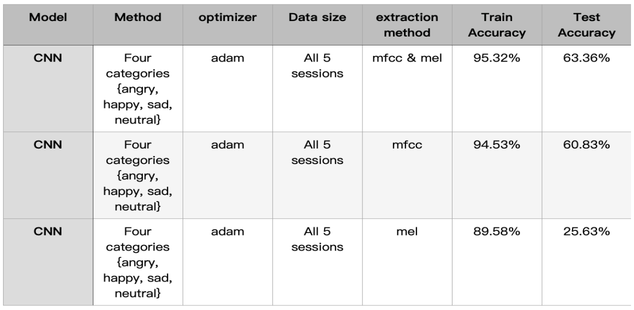
Table 3

Therefore, we try to use CNN model to analysis speech emotion recognition, and the Table 4 & 5 show that the train accuracy and test accuracy by CNN model by different optimizer.



Model	Method	optimizer	Data size	extraction method	Train Accuracy	Test Accuracy
CNN	Four categories {angry, happy, sad, neutral}	sgd	All 5 sessions	mfcc & mel	96.34%	65.162%
CNN	Four categories {angry, happy, sad, neutral}	sgd	All 5 sessions	mfcc	96.02%	61.19%
CNN	Four categories {angry, happy, sad, neutral}	sgd	All 5 sessions	mel	96.67%	22.74%

Table 4



Model	Method	optimizer	Data size	extraction method	Train Accuracy	Test Accuracy
CNN	Four categories {angry, happy, sad, neutral}	adam	All 5 sessions	mfcc & mel	95.32%	63.36%
CNN	Four categories {angry, happy, sad, neutral}	adam	All 5 sessions	mfcc	94.53%	60.83%
CNN	Four categories {angry, happy, sad, neutral}	adam	All 5 sessions	mel	89.58%	25.63%

Table 5

According to Table 4 and 5, the results show that the extraction method has significant impact on test accuracy, and sgd is best optimizer in CNN model.

Finally, we present the confusion matrix to further analyze the SER performances of CNN model. According to Fig.1 and 2, It is obvious to show that on IEMOCAP datasets, the *sad* obtains the highest recognition rate, and *happy* obtains the lowest recognition rate. In sessions 1 of IEMOCAP databases, 20% happy samples are misclassified as sad, and 5% happy samples are misclassified as angry. In all 5 sessions of IEMOCAP datasets, 16.67% happy samples are misclassified as sad, 13.89% happy samples are misclassified as angry, and 8.33% happy are misclassified as neutral. Compared to the four categories emotion analysis model of the IEMOCAP database in 3-D Convolutional Recurrent Neural Networks with Attention Model (Fig. 3), our results show a significant improvement.

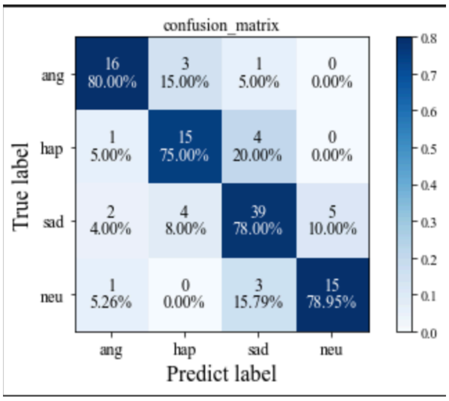


Figure 1- confusion matrix of CNN on session 1 of IEMOCAP

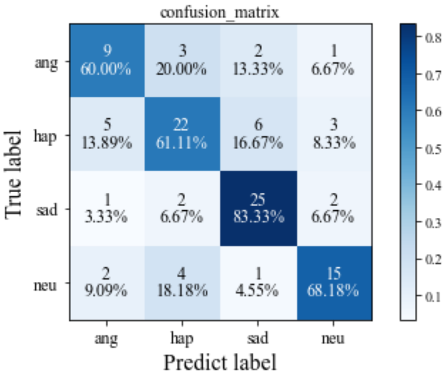
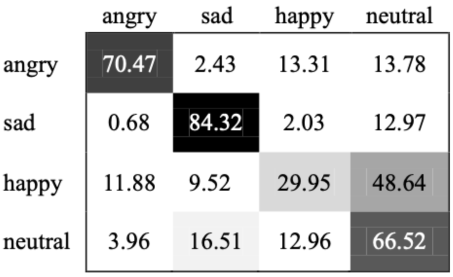


Figure 2- confusion matrix of CNN on all 5 sessions of IEMOCAP



	angry	sad	happy	neutral
angry	70.47	2.43	13.31	13.78
sad	0.68	84.32	2.03	12.97
happy	11.88	9.52	29.95	48.64
neutral	3.96	16.51	12.96	66.52

Figure 3- confusion matrix of 3-D ACRNN