



به نام خدا



دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر
شبکه های عصبی و یادگیری عمیق
مینی پروژه دوم

نام و نام خانوادگی	مهسا تاجیک – مهدیه شاهپوری
شماره دانشجویی	810198181-810198126
تاریخ ارسال گزارش	99/3/20

فهرست گزارش سوالات

قسمت اول : طراحی شبکه های عصبی 3

قسمت دوم (نقصان دادگان) 31

قسمت اول : طراحی شبکه های عصبی

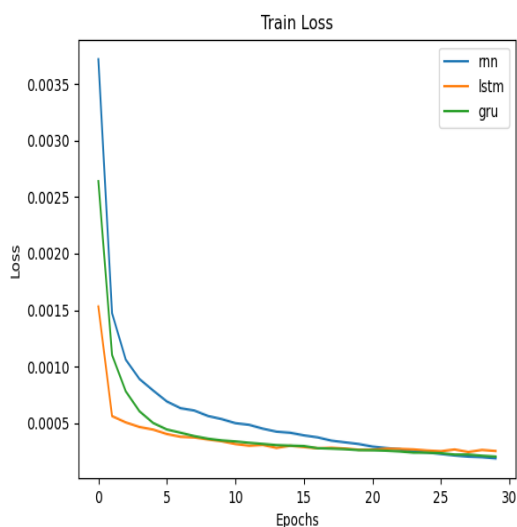
در این قسمت قصد داریم با استفاده از شبکه های بازگشتی ، میزان آلودگی هوا را پیش بینی کنیم به این صورت که با دریافت اطلاعات مربوط به 11 ساعت گذشته ، میزان آلودگی را در ساعت 12 ام و 24 ام پیش بینی می کنیم و به بررسی تفاوت های شبکه های بازگشتی مختلف در انجام اینکار می پردازیم.

دیتاستی که در این تمرین روی آن کار می کنیم دارای ابعاد (43798,8) است که در این قسمت خواسته شده تنها از 15000 سطر ابتدایی داده ها استفاده شود. 12000 سطر برای داده های آموزش و 3000 داده ی بعدی را برای تست جدا میکنیم. داده ها باید بصورت سری های 11 تایی به شبکه وارد شوند بنابراین تابعی بنام data for training نوشتیم که داده ها را با 8 ویژگی شان بصورت 11 تایی از هم جدا می کند و آرایه ای با ابعاد (11,8,11990) برای تحویل به شبکه برمی گرداند. اولین داده از سری زمانی بعدی (ساعت 12ام) و دومین داده از سری زمانی دوتا بعدی (ساعت 24 ام) بعنوان تارگت در آرایه ی y_{train} ذخیره می شوند. Stride برای داده های آموزش و تست برابر با 11 در نظر گرفته شده است. به این نکته توجه داریم که داده ها نرمالیزه شده اند و نیازی به اینکار نیست اما اگر نرمالیزه نشده بودند باید آن ها را نرمالیزه می کردیم.

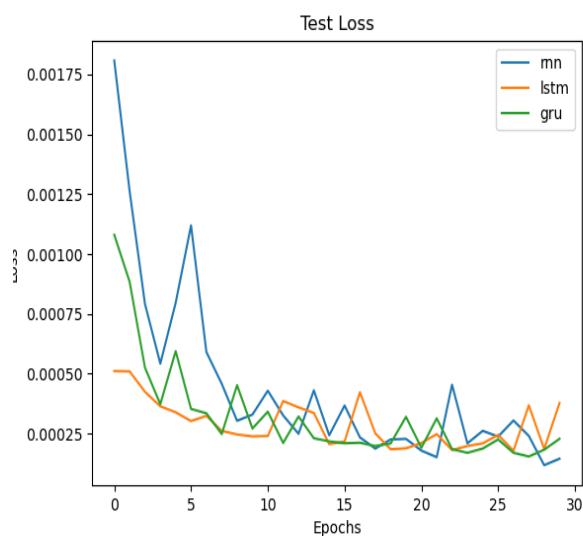
بخش 1و2 :

در این بخش به بررسی نمودار های خطای داده های آموزش و تست و همچنین نمودارهای مقدار پیش بینی و مقدار حقیقی برای سه شبکه ی بازگشتی RNN , LSTM , GRU میپردازیم.

در این قسمت مدل ها با معیار خطای mse و آپتیمایزر RMSPROP کامپایل شده اند. تعداد ایپاک ها برای هر سه شبکه برابر با 30 و batch_Size برابر با 30 در نظر گرفته شده است و 0.2 از داده های آموزش برای ارزیابی جدا شده اند.



نمودار 1-1- خطای داده های آموزش برحسب هر اپیک



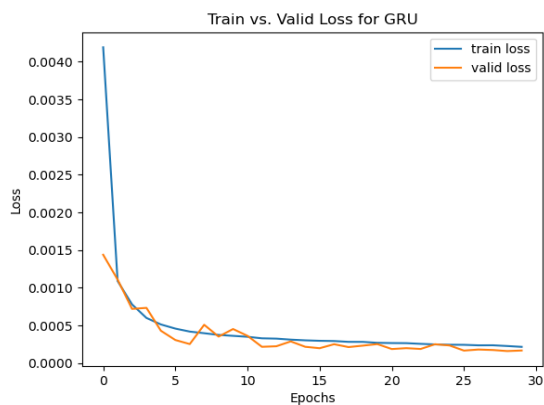
نمودار 1-2- خطای داده های تست برحسب هر اپیک

همانطور که در نمودار های 1-1 و 1-2 میبینیم مقدار خطای داده ها برای شبکه ی lstm از دو شبکه ی دیگر کمتر بوده و rnn خطای بیشتری از هردوی آن ها دارد.

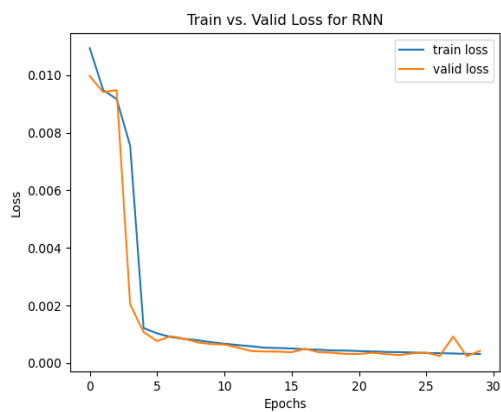
```
learning time for RNN : 29.151723631000003
learning time for LSTM : 122.578851318
learning time for GRU : 99.19922801199996
```

شکل 1-1- زمان آموزش برای سه شبکه

در شکل 1-1 مشاهده می کنیم که زمان آموزش شبکه برای lstm از همه بیشتر و rnn بسیار کمتر از بقیه است.



نمودار 1-4- خطای داده های آموزش و ارزیابی برای
GRU



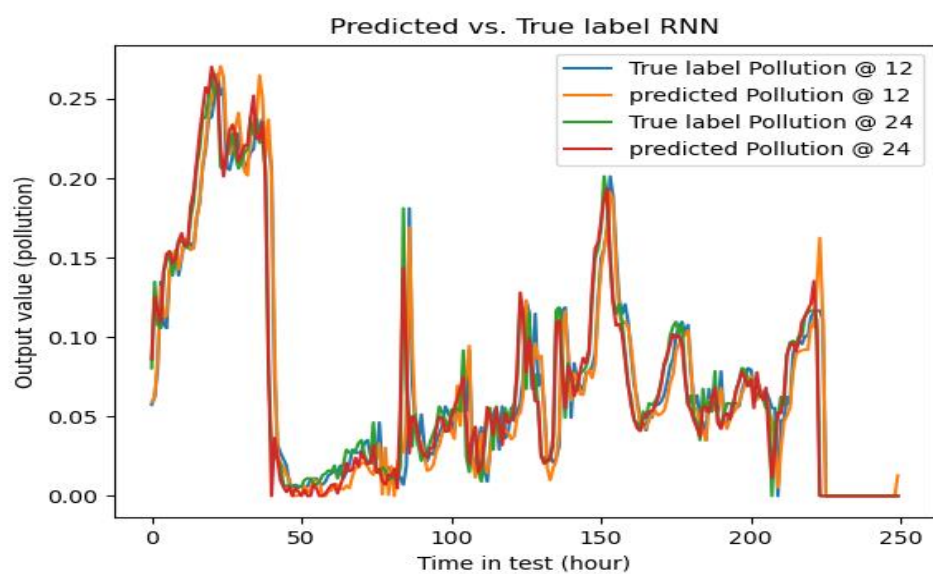
نمودار 1-3- خطای داده های آموزش و ارزیابی برای
RNN



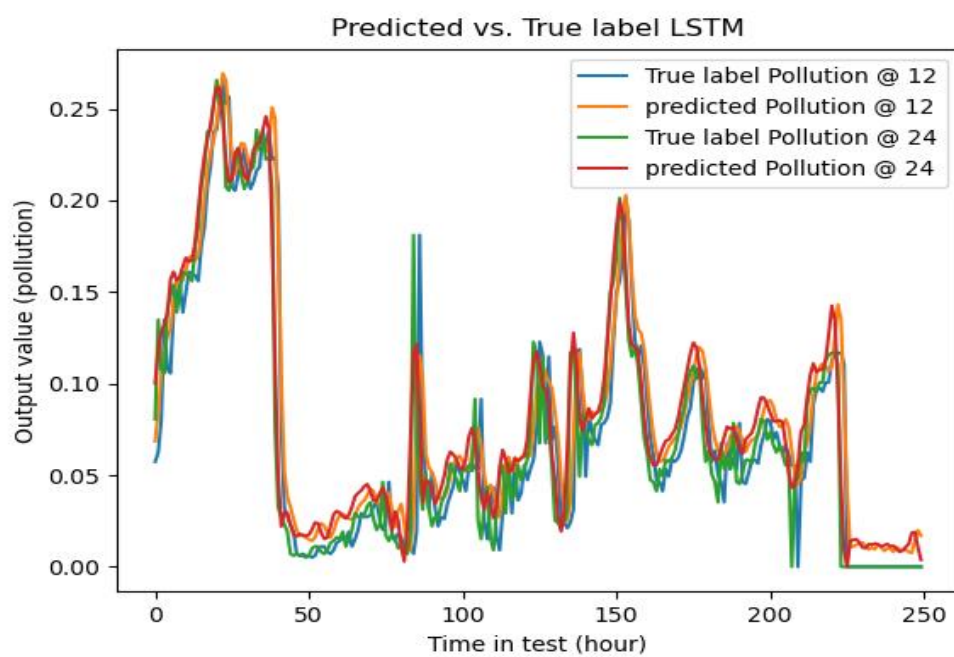
نمودار 1-5- خطای داده های آموزش و ارزیابی برای
LSTM

در نمودارهای 1-3 و 1-4 و 1-5 خطای داده های آموزش و ارزیابی را برای هر شبکه بصورت جدا

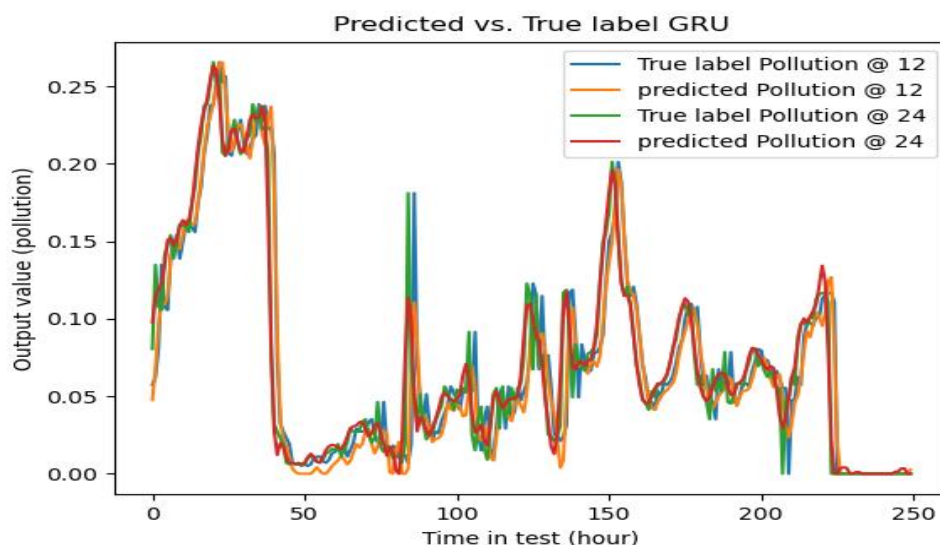
میبینیم.



نمودار 6-1- مقدار حقیقی و پیش بینی برای شبکه RNN



نمودار 7-1- مقدار حقیقی و پیش بینی برای شبکه LSTM

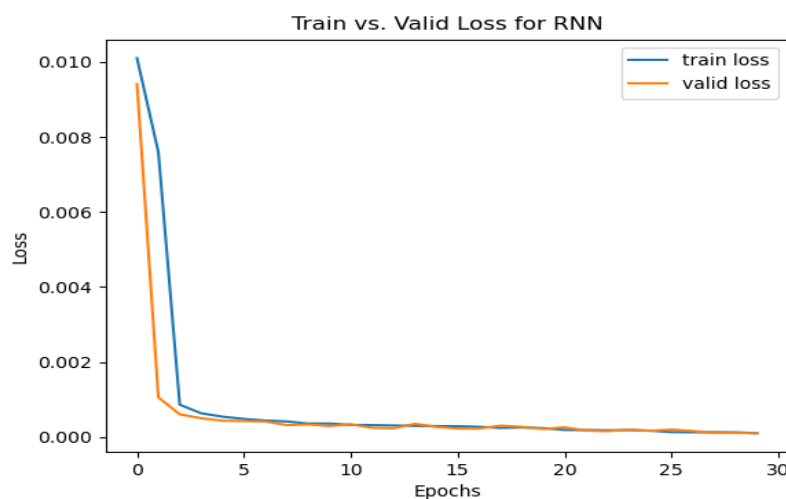


نمودار 1-8- مقدار حقیقی و پیش بینی برای شبکه GRU

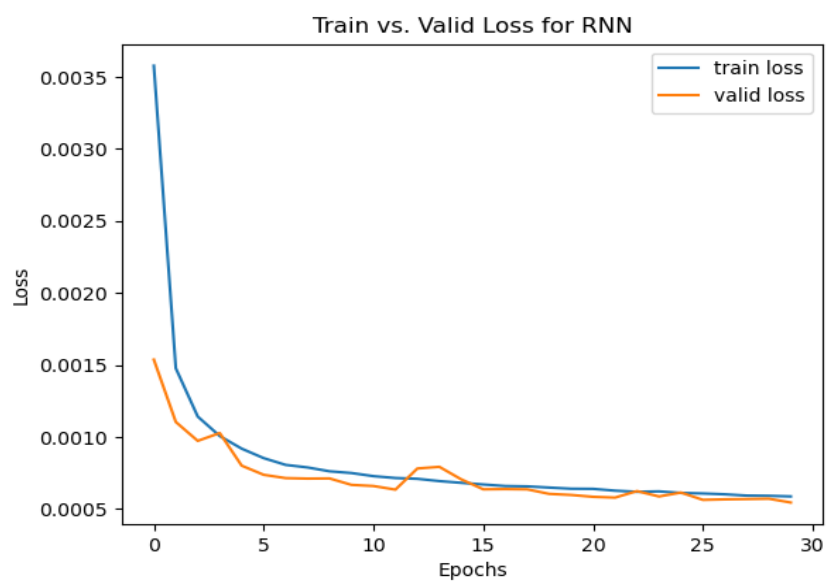
در نمودار های 1-6 ، 1-7 و 1-8 مقادیر حقیقی و پیش بینی برای سه شبکه آمده است و برای اینکه نمودار واضح تر باشد تنها 250 داده ی اول را نشان داده ایم و مشاهده می کنیم شبکه ی gru پیش بینی نسبتاً بهتری انجام داده ولی هر سه شبکه بسیار نزدیک به هم هستند.

بخش 3:

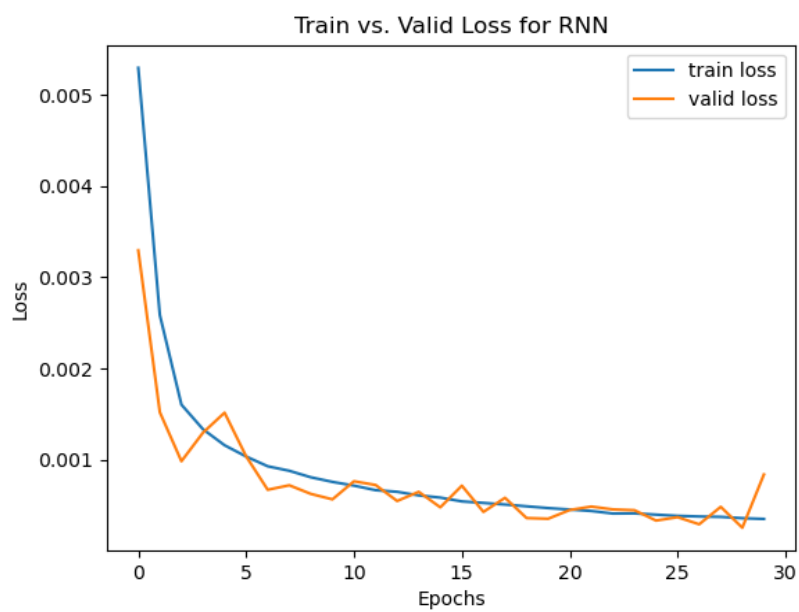
در این بخش به بررسی هر سه شبکه با معیارهای خطای متفاوت و روش های بهینه سازی متفاوت می پردازیم.



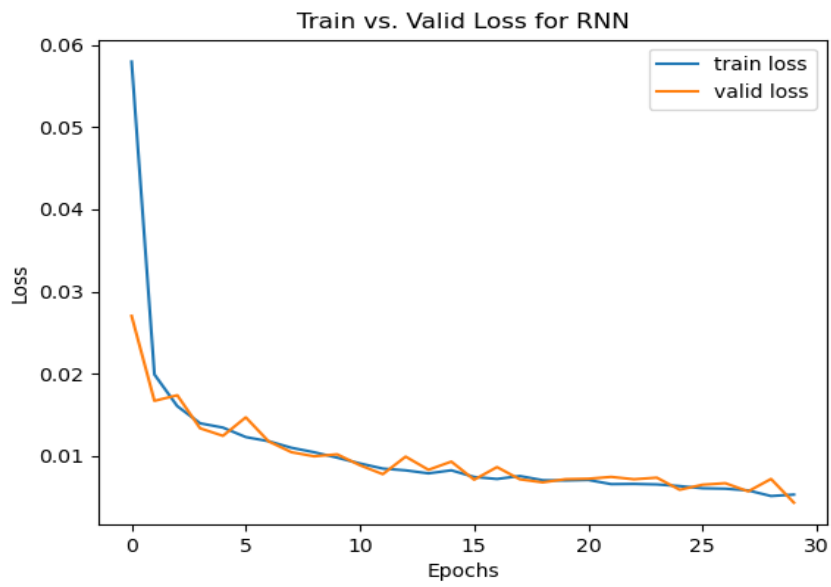
Rnn-adam-mse



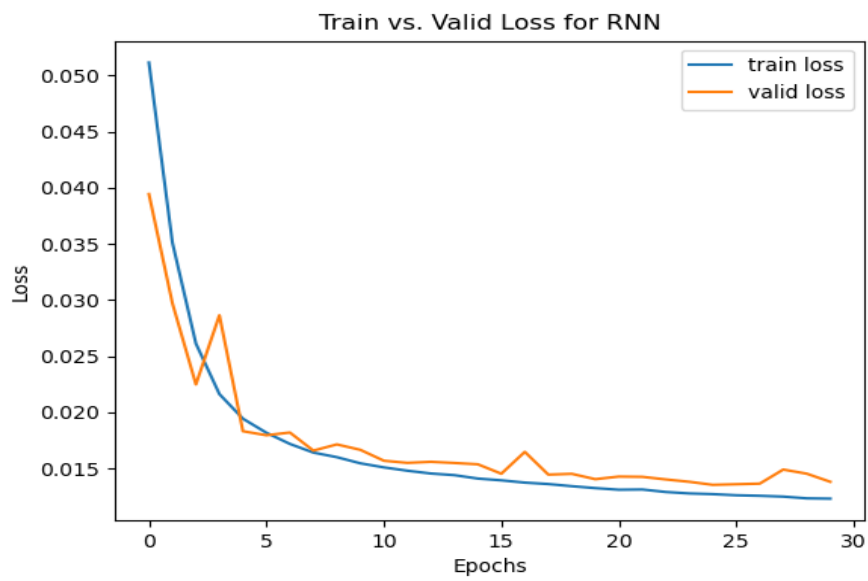
Rnn-adagrad-mse



Rnn-rmsprop-mse



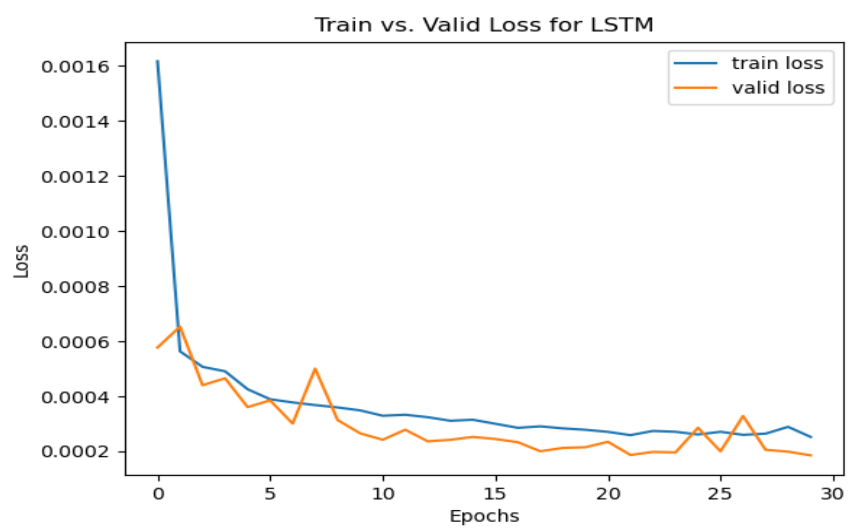
Rnn-adam-mae



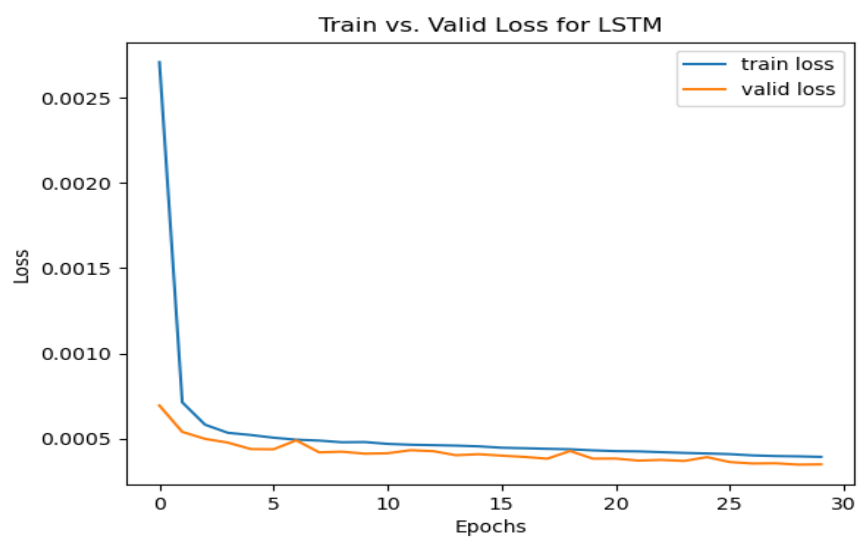
Rnn-adagrad-mae



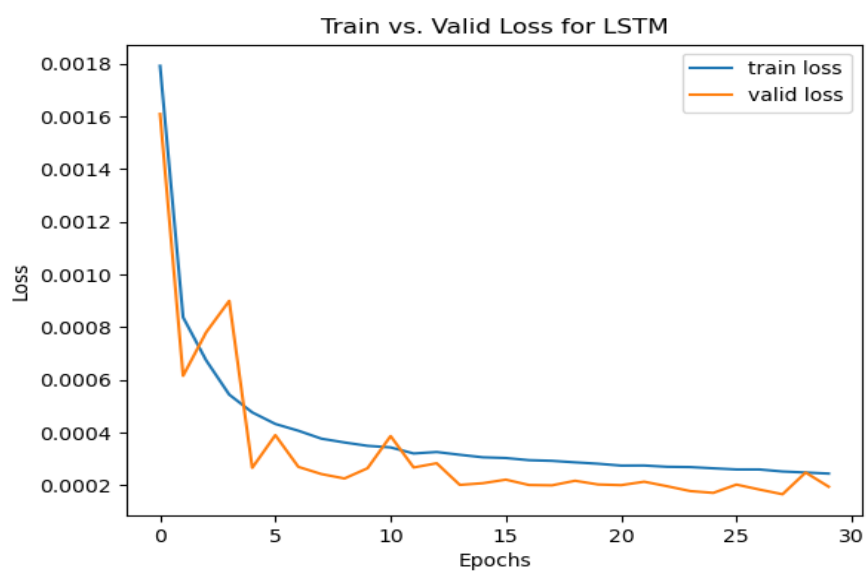
Rnn-rmsprop-mae



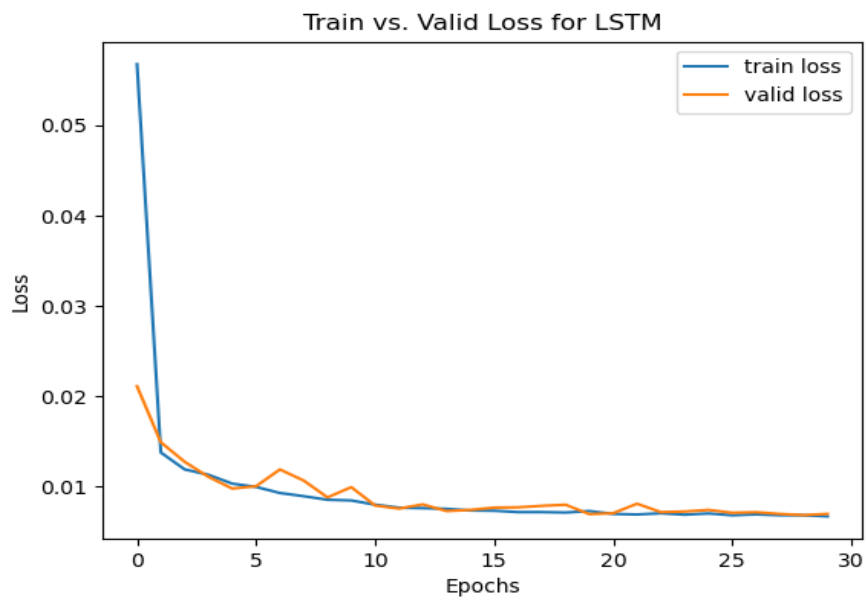
Lstm-adam-mse



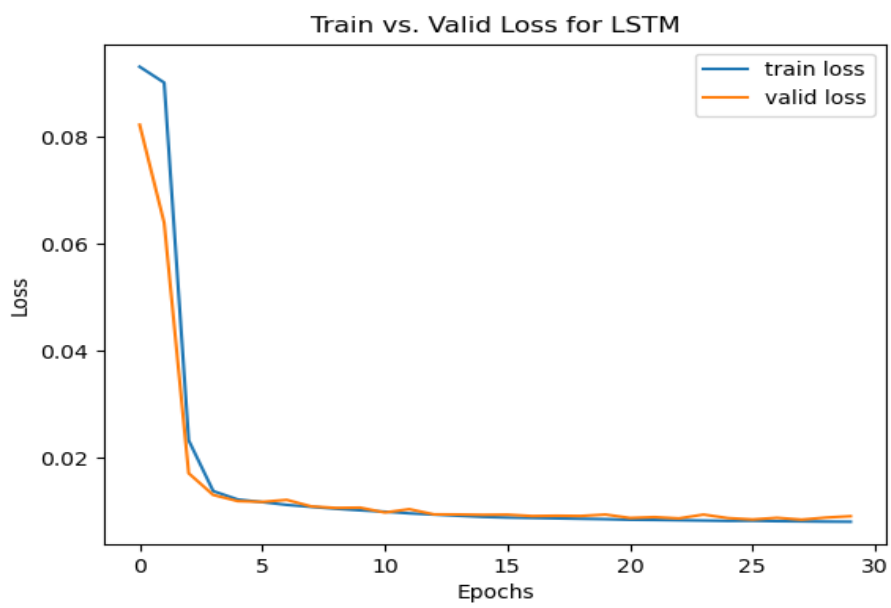
Lstm-adagrad-mse



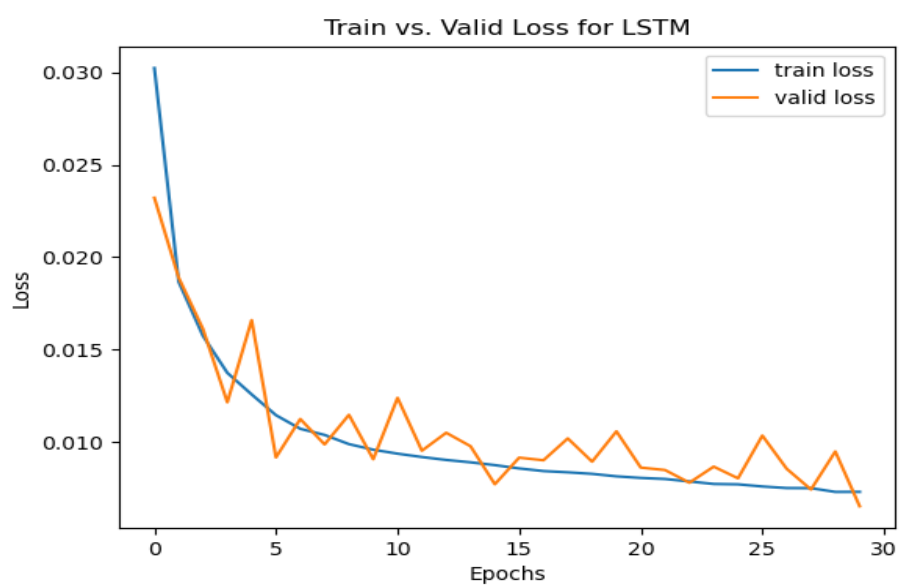
Lstm-rmsprop-mse



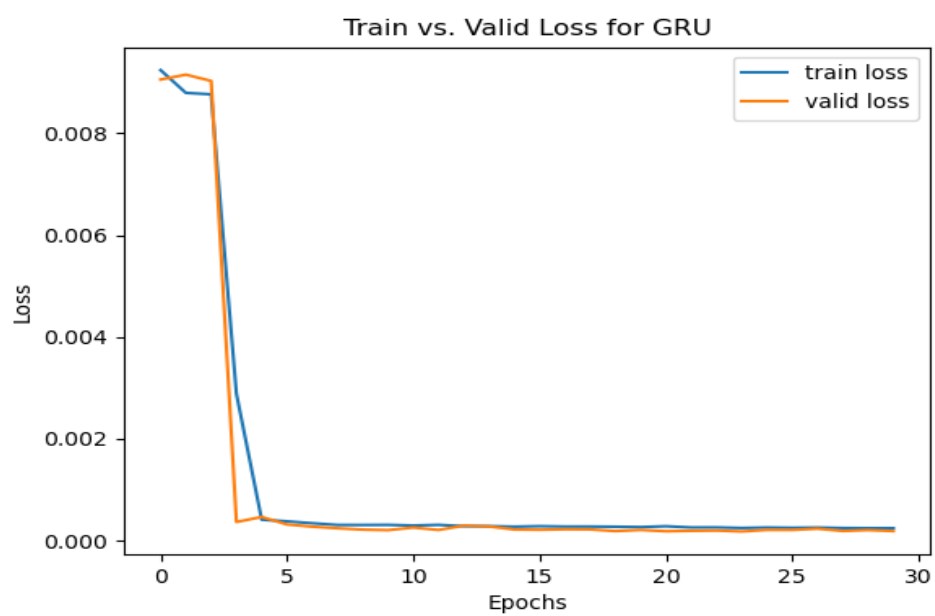
Lstm-adam-mae



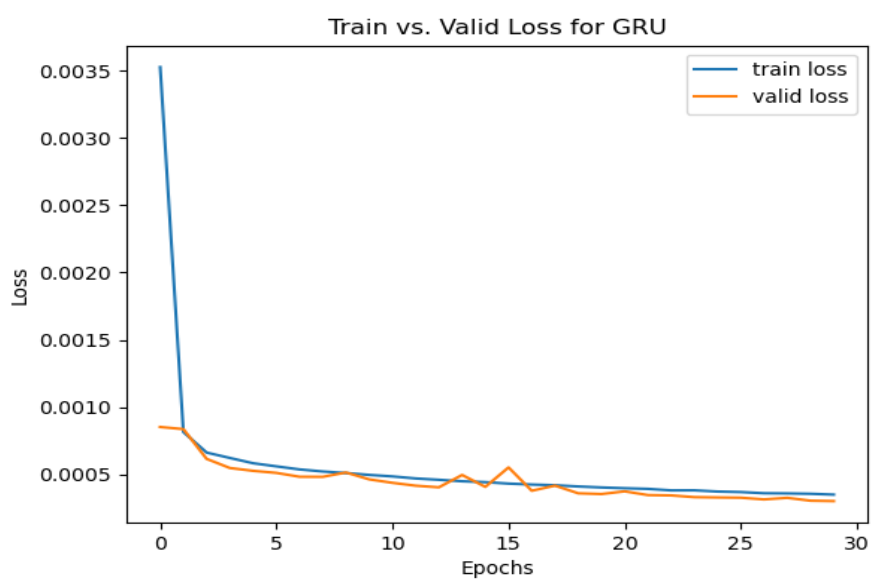
Lstm-adagrad-mae



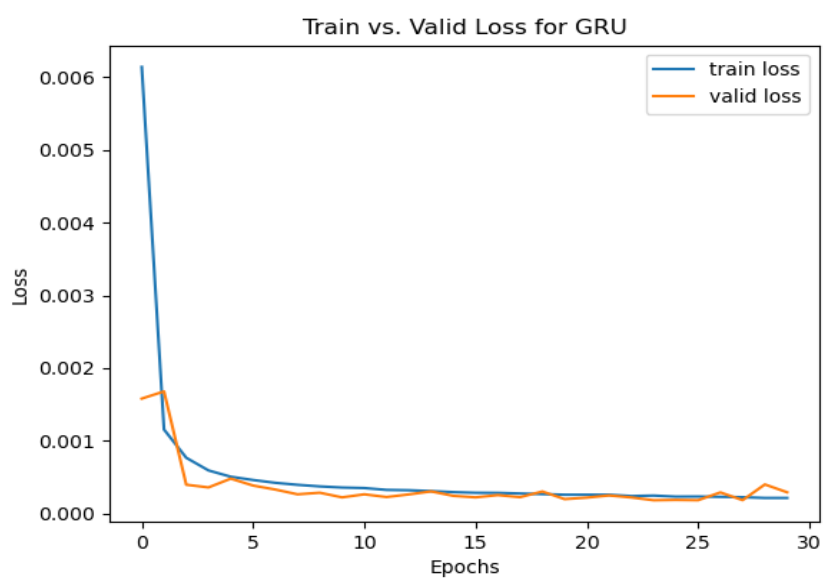
Lstm-rmsprop-mae



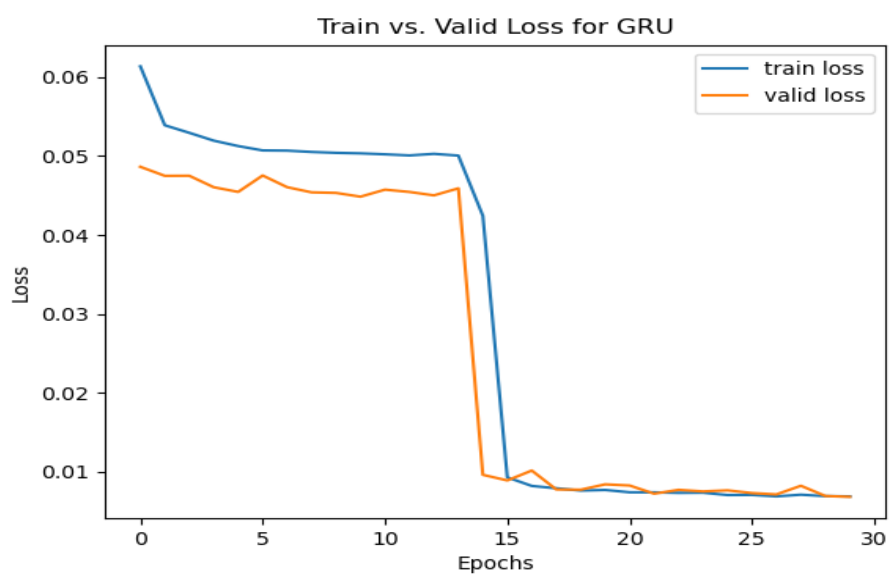
Gru-adam-mse



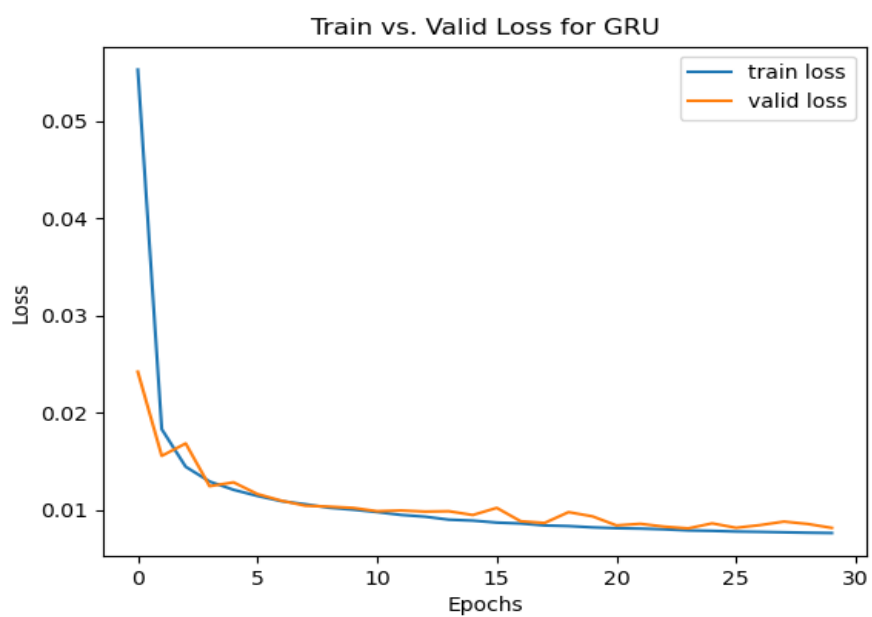
Gru-adagrad-mse



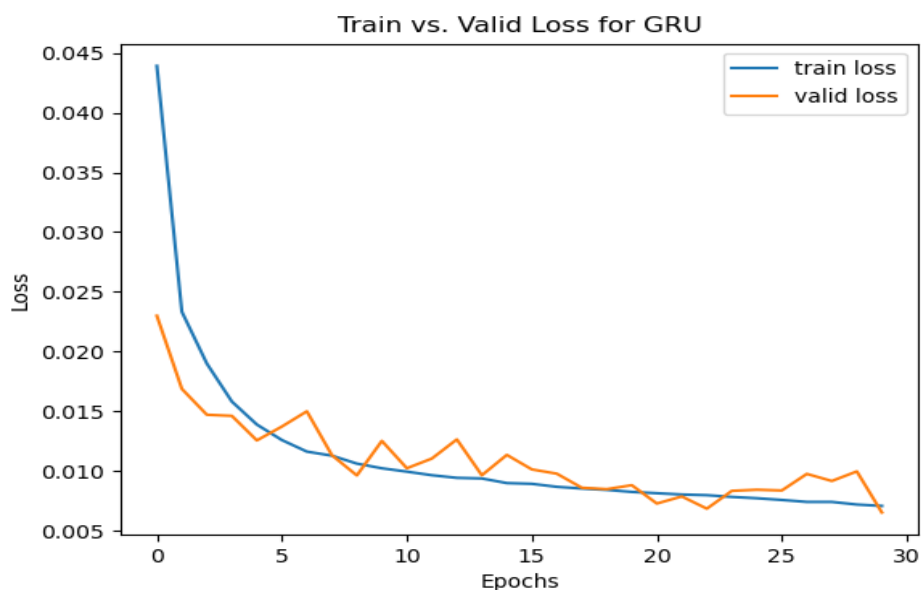
Gru-rmsprop-mse



Gru-adam-mae



Gru-adagrad-mae



Gru-rmsprop-mae

همانطور که در نمودارهای بالا و شکل های 1-2 تا 1-7 میبینیم مقدار خطا برای شبکه ی **rnn** با بهینه ساز **adagrad** و معیار خطای **mse** کمترین میزان است. برای شبکه ی **lstm** بهینه ساز های **adagrad** و **rmsprop** با معیار خطای **mse** بهتر از بقیه عمل می کنند و برای شبکه ی **gru** بهینه ساز های **adam** و **rmsprop** با معیار خطای **mse** کمترین خطا را در بین کلیه ی شبکه ها دارند. در تمامی حالات معیار خطای **mse** نتیجه ی بهتری داشته بجز زمانیکه شبکه ی **rnn** با بهینه ساز **adam** کامپایل میشود تنها در این حالت معیار **mae** جواب بهتری داده است. می توان گفت از لحاظ میزان خطا شبکه ی **gru** کمترین مقدار خطا را دارد.

میزان دقیق خطا و همچنین زمان آموزش شبکه برای تمام حالت ها را در شکل های 1-2 تا 1-7 میبینیم:

```
learning time for RNN-adam-MSE : 27.78026234
learning time for LSTM-adam-MSE : 113.831367702
learning time for GRU-adam-MSE : 103.39212529
```

```
RNN-adam-MSE :
train loss: 0.008540163200872381
test loss: 0.006977534281153366
```



```
LSTM-adam-MSE Loss :  
train loss: 0.00022046927957164556  
test loss: 0.00013245233413638753  
Model: "Sequential(2)"
```

```
GRU-adam-MSE Loss :  
train loss: 0.0002049748688953468  
test loss: 0.000130564310773078
```

شکل 2-1- خطا و زمان آموزش برای سه شبکه با بهینه ساز adam و معیار خطای mse

```
learning time for RNN-adagrad-MSE : 26.04742227  
learning time for LSTM-adagrad-MSE : 122.615367611  
learning time for GRU-adagrad-MSE : 113.066831334
```

```
RNN-adagrad-MSE :  
train loss: 0.0002951813219436336  
test loss: 0.00019473316533914238
```

```
LSTM-adagrad-MSE Loss :  
train loss: 0.0003190957476514819  
test loss: 0.00017457331089562738
```

```
GRU-adagrad-MSE Loss :  
train loss: 0.0003309885250081811  
test loss: 0.00017706743620330605
```

شکل 3-1- خطا و زمان آموزش برای سه شبکه با بهینه ساز adagrad و معیار خطای mse

```
learning time for RNN-RMSprop-MSE : 33.917271765  
learning time for LSTM-RMSprop-MSE : 134.494332066  
learning time for GRU-RMSprop-MSE : 109.73750650400001
```

```
RNN-RMSprop-MSE :  
train loss: 0.008729875112440376  
test loss: 0.007185697022912694
```

```
LSTM-RMSprop-MSE Loss :  
train loss: 0.00025045830560587904  
test loss: 0.00019237244132120493
```

```
GRU-RMSprop-MSE Loss :  
train loss: 0.00020216152844800114  
test loss: 0.0001264889835970295
```

شکل 4-1- خطا و زمان آموزش برای سه شبکه با بهینه ساز rmsprop و معیار خطای mse

```
learning time for RNN-adam-MAE : 36.849865664000006
learning time for LSTM-adam-MAE : 131.991191767
learning time for GRU-adam-MAE : 106.67267067200001
```

```
RNN-adam-MAE :
train loss: 0.004174686074365642
test loss: 0.0038883727258787705
```

```
LSTM-adam-MAE Loss :
train loss: 0.006162583962350064
test loss: 0.0062749750585136405
```

```
GRU-adam-MAE Loss :
train loss: 0.007018617178011279
test loss: 0.006450011612953351
```

شکل 5-1- خطا و زمان آموزش برای سه شبکه با بهینه ساز adam و معیار خطای mae

```
learning time for RNN-adagrad-MAE : 30.153239897
learning time for LSTM-adagrad-MAE : 120.854589337
learning time for GRU-adagrad-MAE : 104.261191979000002
```

```
RNN-adagrad-MAE :
train loss: 0.013260373780604847
test loss: 0.012843337000961288
```

```
LSTM-adagrad-MAE Loss :
train loss: 0.00801380208329926
test loss: 0.007534634453527953
```

```
GRU-adagrad-MAE Loss :
train loss: 0.0903832473047233
test loss: 0.09263877263510706
```

شکل 6-1- خطا و زمان آموزش برای سه شبکه با بهینه ساز adagrad و معیار خطای mae

```
learning time for RNN-RMSprop-MAE : 29.942773127000002
learning time for LSTM-RMSprop-MAE : 115.090278376
learning time for GRU-RMSprop-MAE : 105.334326807
```

```
RNN-RMSprop-MAE :
train loss: 0.012661671135776533
test loss: 0.011787554120215087
Model: "sequential_2"
```

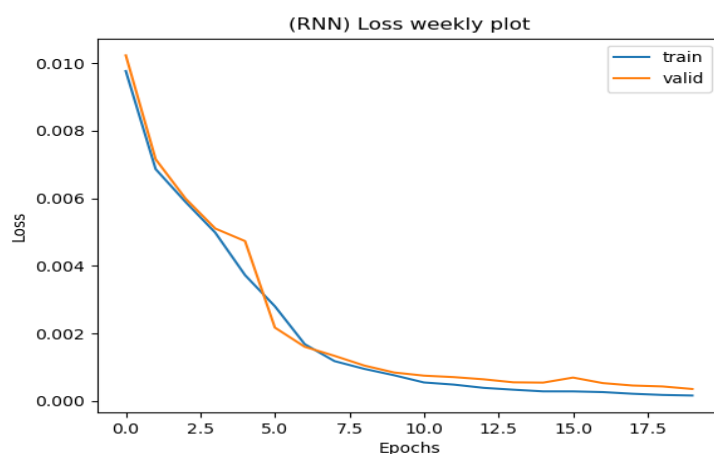
```
LSTM-RMSprop-MAE Loss :
train loss: 0.006268860396616055
test loss: 0.005869376470014004
```

```
GRU-RMSprop-MAE Loss :
train loss: 0.007501427315523517
test loss: 0.0071264082727872805
```

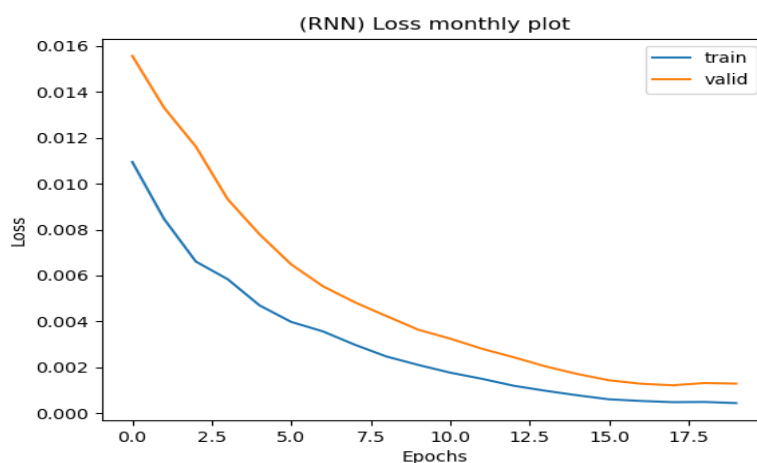
شکل 7-1- خطا و زمان آموزش برای سه شبکه با بهینه ساز rmsprop و معیار خطای mae همانطور که در شکل های 2-1 تا 7-1 مشاهده می شود زمان آموزش شبکه ی rnn در تمام حالت ها بسیار کمتر از دو شبکه ی دیگر است و دو شبکه ی دیگر زمان اجرای نسبتا نزدیکی به یکدیگر دارند.

بخش 4:

در این قسمت می خواهیم با تولید یک عدد بصورت رندوم برای یک ساعت مشخص از هر روز با دریافت داده های 6 روز گذشته میزان آلودگی در روز 7 ام در ساعتی مشخص را پیش بینی کنیم و با تولید یک روز و ساعت رندوم از هفته، با دریافت داده های 3 هفته گذشته ، میزان آلودگی در ساعت و روز مشخصی از هفته ی 4 ام را پیش بینی کنیم. از آنجاییکه نیاز به داده های بیشتری داریم از تمام داده های دیتاست برای این قسمت استفاده میکنیم و 30000 داده ی اول را برای آموزش و بقیه ی داده ها را برای تست جدا میکنیم. 0.2 از داده های آموزش را برای ارزیابی جدا می کنیم. داده ها برای 20 اپیک آموزش میبینند.

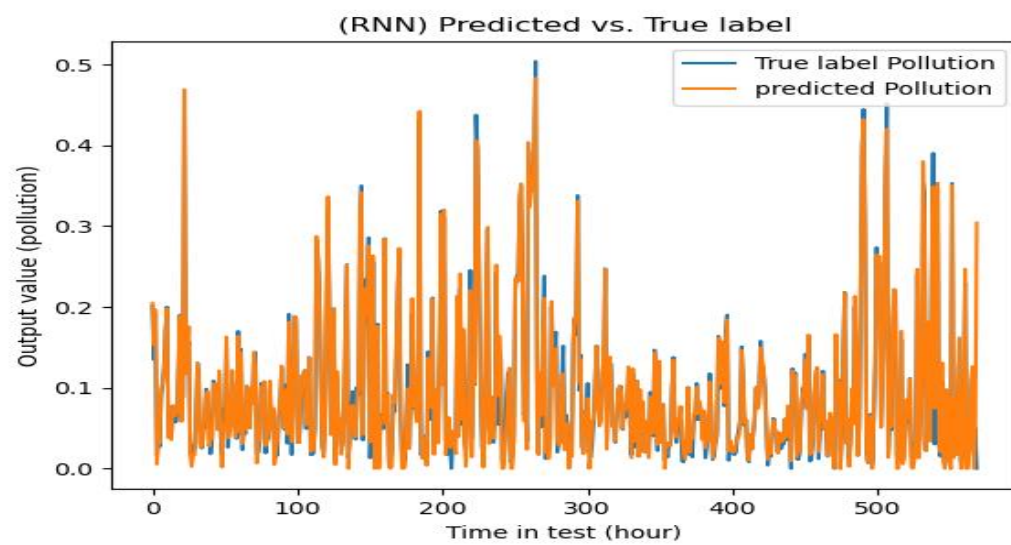


نمودار خطا برای پیش بینی هفتگی

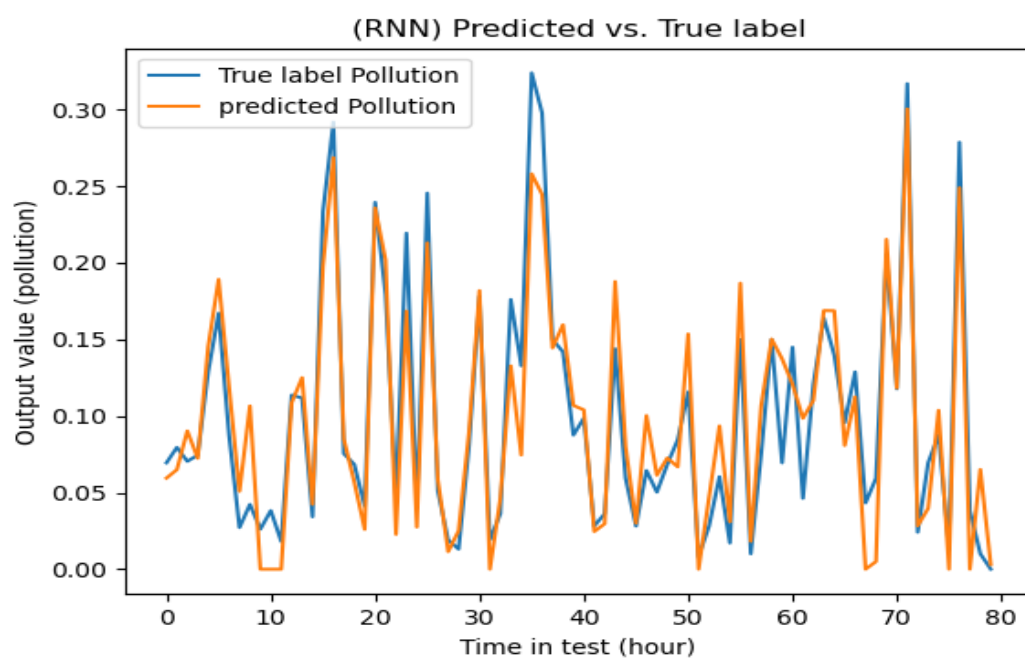


نمودار خطا برای پیش بینی ماهیانه

همانطور که انتظار داشتیم میزان خطا برای پیش بینی هفتگی کمتر از ماهیانه است علتش هم اینست که برای پیش بینی های هفتگی از 6 داده استفاده میکنیم ولی برای پیش بینی های ماهیانه تنها از 3 داده استفاده می کنیم و علت دیگر اینکه تعداد داده های کلی برای پیش بینی های هفتگی نسبت به ماهیانه بیشتر است و باعث آموزش بهتر شبکه می شود. این موضوع را در نمودارهای مقایسه حقیقی و پیش بینی شده هم میتوانیم ببینیم.



نمودار مقدار حقیقی و پیش بینی برای سری زمانی هفتگی



نمودار مقدار حقیقی و پیش بینی برای سری زمانی ماهیانه

```
RNN weekly loss :  
train loss: 0.00044269505930301476, train acc: 0.01682845503091812  
test loss: 0.000619631327945667, test acc: 0.019237995147705078  
weekly learning time for RNN : 2.458721742
```

شکل 8-1- مقادیر خطا و زمان آموزش شبکه برای سری زمانی هفتگی

```
RNN monthly loss :  
train loss: 0.0020501214634669196, train acc: 0.03142104297876358  
test loss: 0.004112319927662611, test acc: 0.0403759740293026  
monthly learning time for RNN : 1.137560112
```

شکل 9-1- مقادیر خطا و زمان آموزش شبکه برای سری زمانی ماهیانه

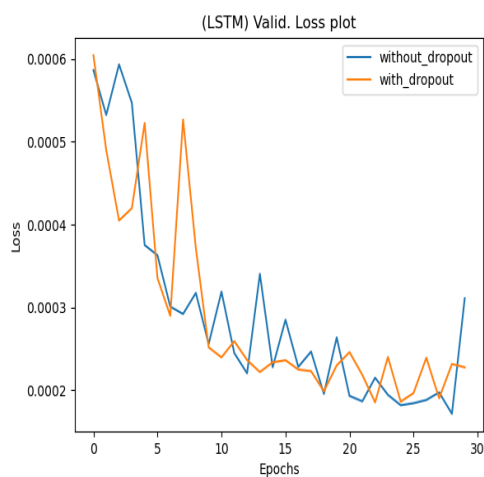
زمان آموزش شبکه برای داده های ماهیانه کمتر است چون تعداد داده ها در این حالت کمتر است.

بخش 5:

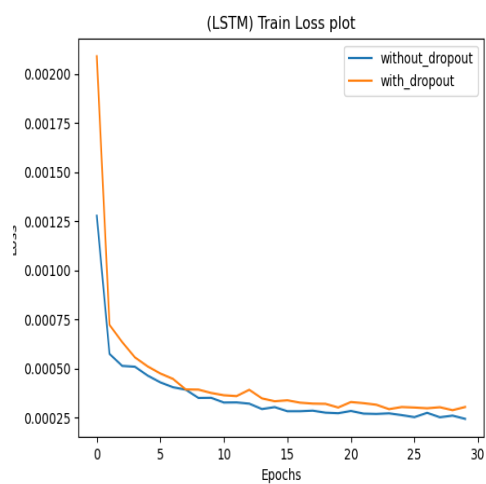
در این بخش می خواهیم عملکرد شبکه را با و بدون لایه dropout بررسی کنیم.

شبکه lstm را برای بررسی این بخش انتخاب می کنیم و مدل را با معیار خطای mse و بهینه ساز adam یکبار با و یکبار بدون لایه dropout کامپایل می کنیم . تعداد ایپاک ها برابر 30 تاست.

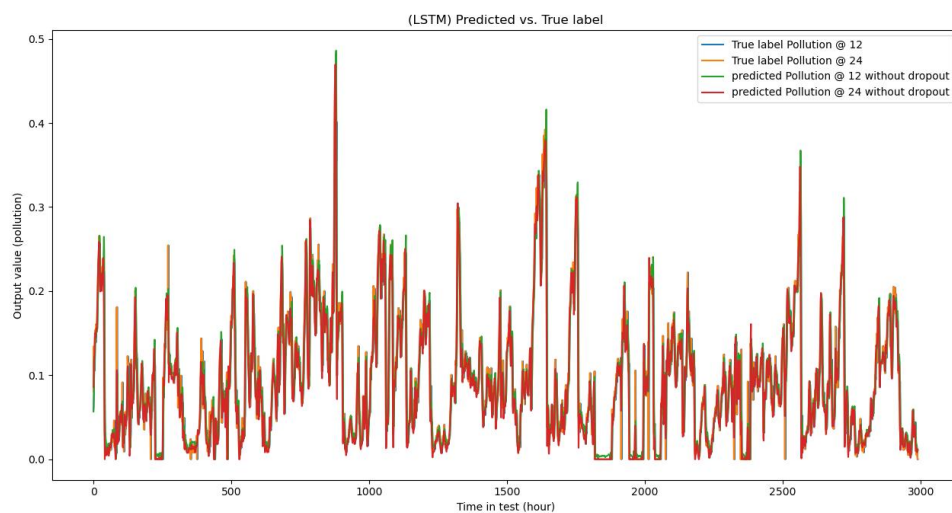
همانطور که می دانیم لایه dropout از overfit شدن جلوگیری می کند و میزان آن را کاهش میدهد بنابراین انتظار داریم مقدار خطای ارزیابی کمی بیشتر شود و مقادیر پیش بینی به مقادیر واقعی نزدیکتر شوند که در نمودارهای خطایی که در زیر آمده میبینیم که میزان خطای داده های ارزیابی اندکی از خطای داده های آموزش بیشتر است.



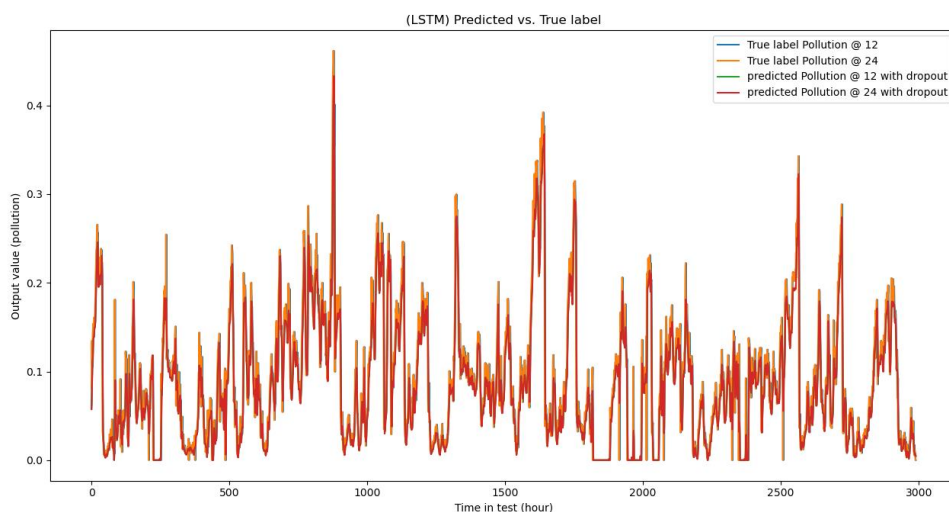
خطای داده ارزیابی با و بدون لایه dropout



خطای داده آموزش با و بدون لایه dropout



نمودار مقادیر پیش بینی و واقعی بدون لایه dropout

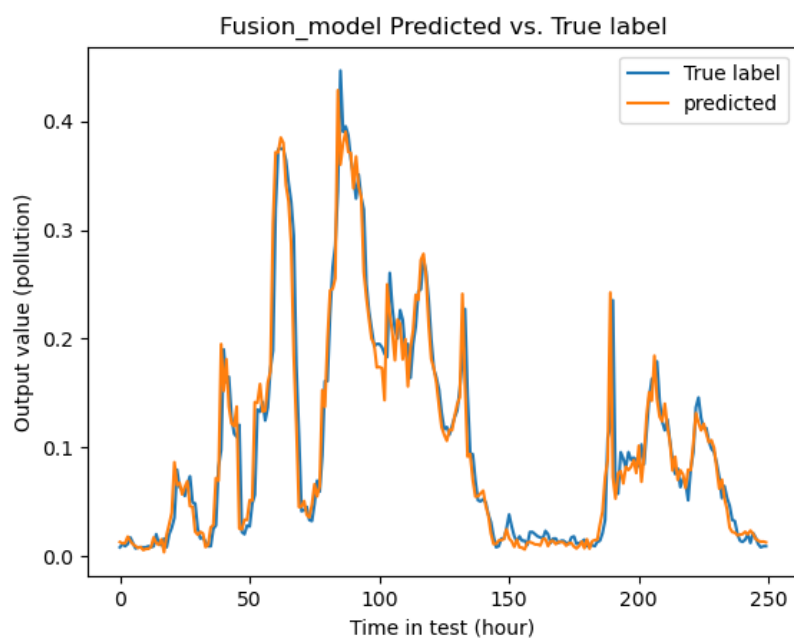


نمودار مقادیر پیش بینی و واقعی با لایه dropout

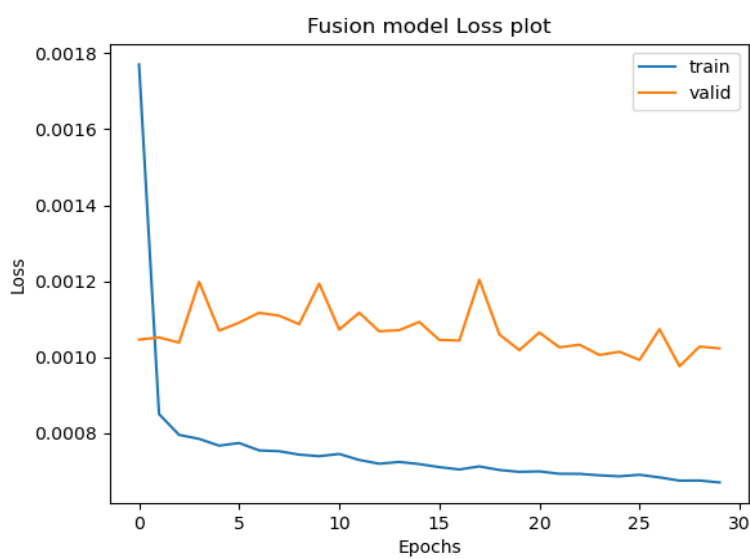
بخش 6:

در این بخش خواسته شده بهترین شبکه ی بخش های قبل را انتخاب کنیم و از آن 3 مدل ساخته و با هم ترکیب کنیم و سه سری زمانی روزانه ، هفتگی و ماهیانه را به مدل ترکیبی بدهیم و خروجی را بررسی کنیم. بنابراین در این قسمت نحوه ی ساختن داده ها متفاوت با قبل بوده و باید برای پیش بینی یک ساعت مشخصی از یک روز مشخص از هفته ، داده های 11 ساعت قبل ، 6 روز قبل و 3 هفته قبل از آن را دریافت کنیم و آلودگی آن ساعت را پیش بینی کنیم. این 3 دسته داده هر کدام یک آرایه می شود و این 3 آرایه را بعنوان داده های آموزش به شبکه ی ترکیبی می دهیم تا مدل فیت شود.

بهترین شبکه در بخش 3 دیدیم شبکه ی GRU با معیار خطای mse و بهینه ساز adam و rmsprop بود که در این قسمت ما بهینه ساز adam را انتخاب میکنیم. نمودار مقدار پیش بینی و حقیقی را در زیر برای 250 داده ی اول میبینیم که پیش بینی خوبی انجام شده است.



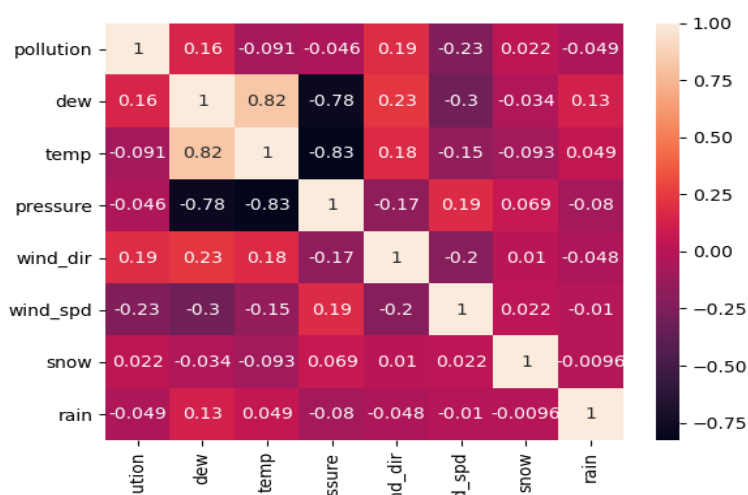
نمودار مقدار پیش بینی شده و مقدار واقعی برای شبکه فیوژن



نمودار خطای داده های آموزش و ارزیابی برای شبکه فیوژن

بخش 7:

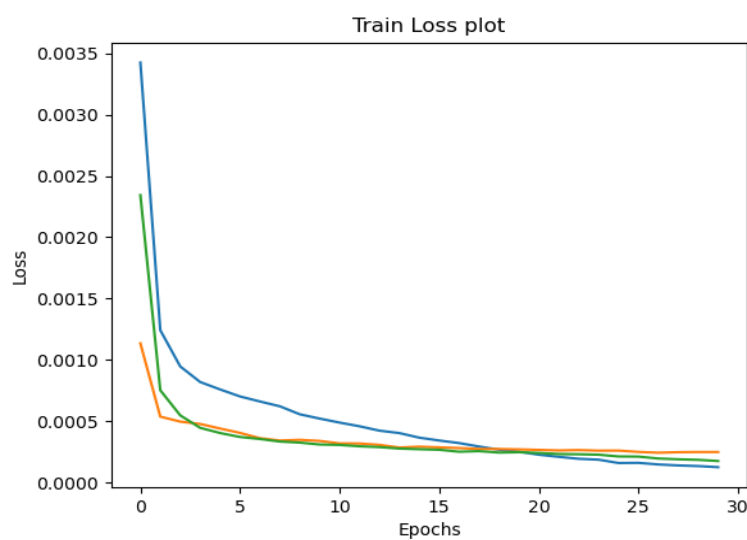
در این بخش خواسته شده دو ستونی را انتخاب کنیم که بیشترین تاثیر را در پیش بینی آلودگی داشته باشند. برای اینکار نیاز داریم تا ویژگی هایی که ارتباط بیشتری با آلودگی دارند انتخاب کنیم و ویژگی های دیگر را حذف کنیم. حذف ویژگی هایی که تاثیر کمتری در پیش بینی ما دارند باعث میشود دقت پیش بینی ها بالاتر رود. از جمله روش هایی که می توان برای انتخاب بهترین ویژگی ها استفاده کرد ، feature selection و backward elimination هستند ولی این روش ها بسیار زمان بر هستند و در اینجا از correlation بین ویژگی ها استفاده می کنیم و با رسم ماتریس correlation دو ویژگی ای که بیشترین correlation را با آلودگی دارند و خودشان correlation بالایی باهم ندارند ، انتخاب می کنیم. بیشترین همبستگی با آلودگی را در ویژگی های wind_dir و wind_spd می بینیم.



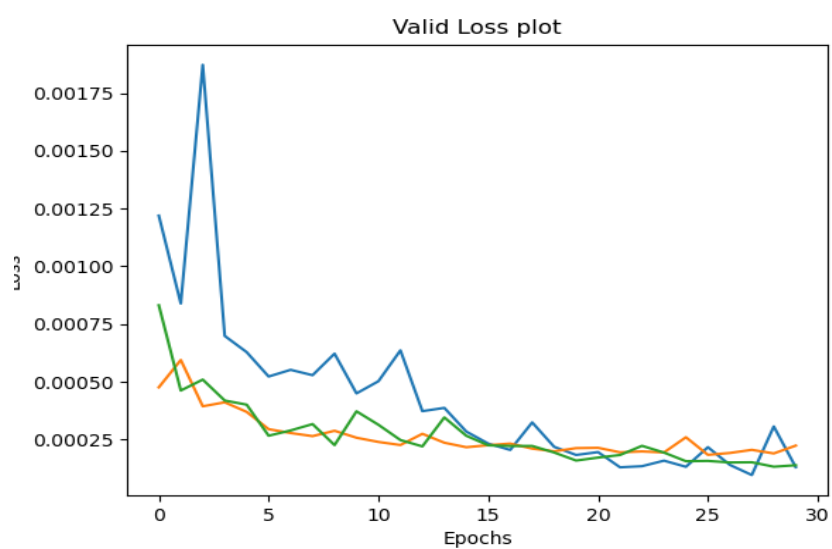
بخش 8 :

در این بخش خواسته شده میزان آلودگی را تنها با استفاده از 3 ستون آلودگی ، سرعت و جهت باد که در قسمت قبل دیدیم بیشترین همبستگی را با آلودگی دارند، برای هر سه شبکه پیش بینی کنیم. از معیار خطای mse و بهینه ساز adam استفاده می کنیم . تعداد اپاک ها و batch_size 30 در نظر گرفته شده است.

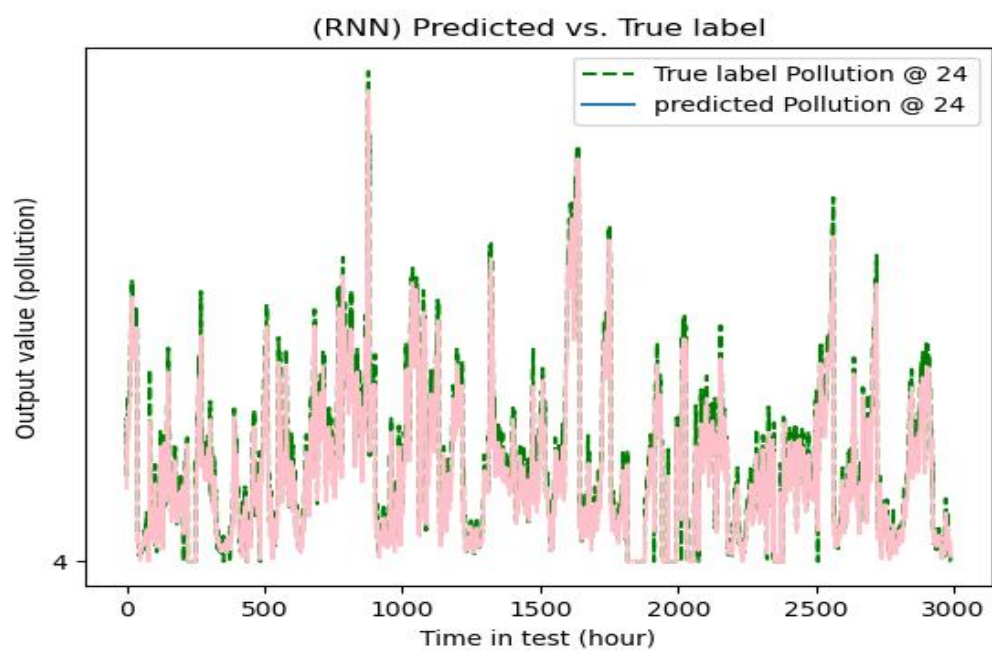
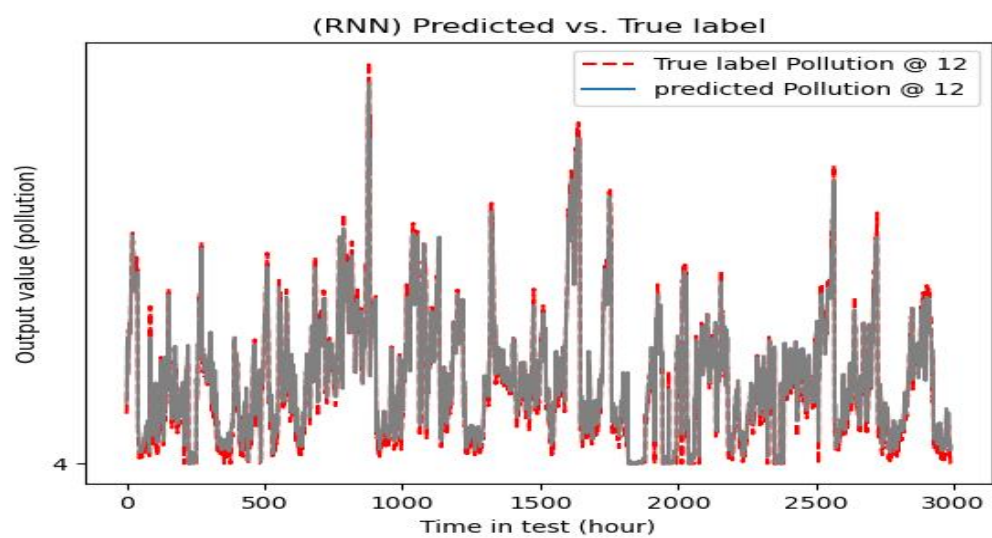
همانطور که می بینیم و انتظار داشتیم با حذف ویژگی های غیرمفید خطای هر سه شبکه برای داده های آموزش و ارزیابی کاهش داشته است.



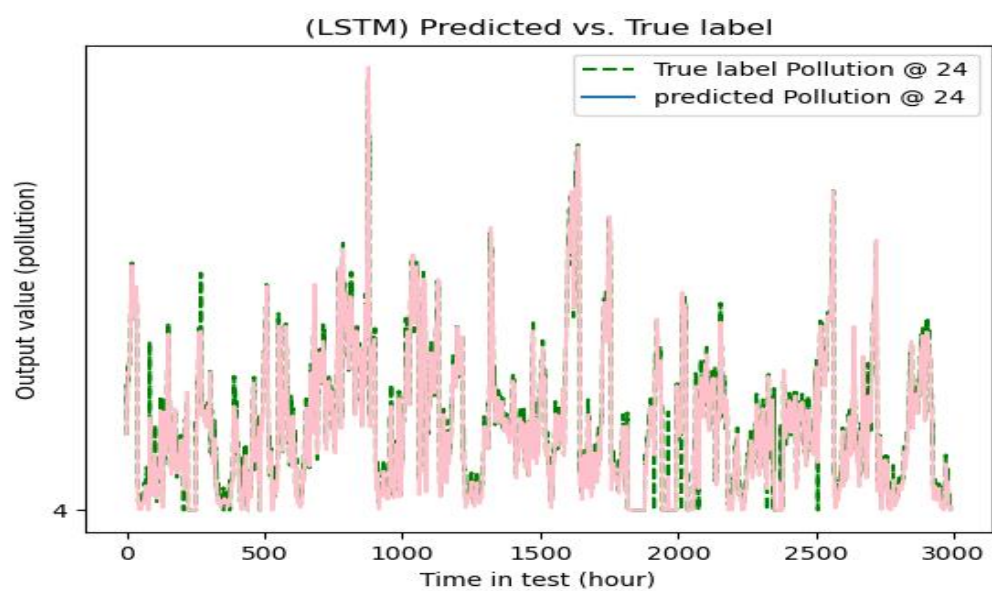
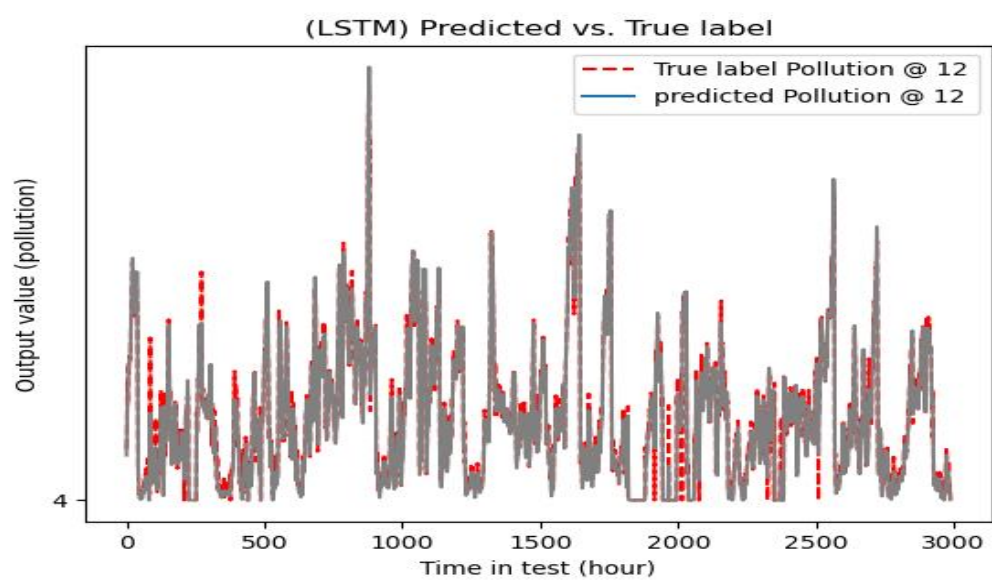
نمودار خطای داده ی آموزش برای 3 شبکه



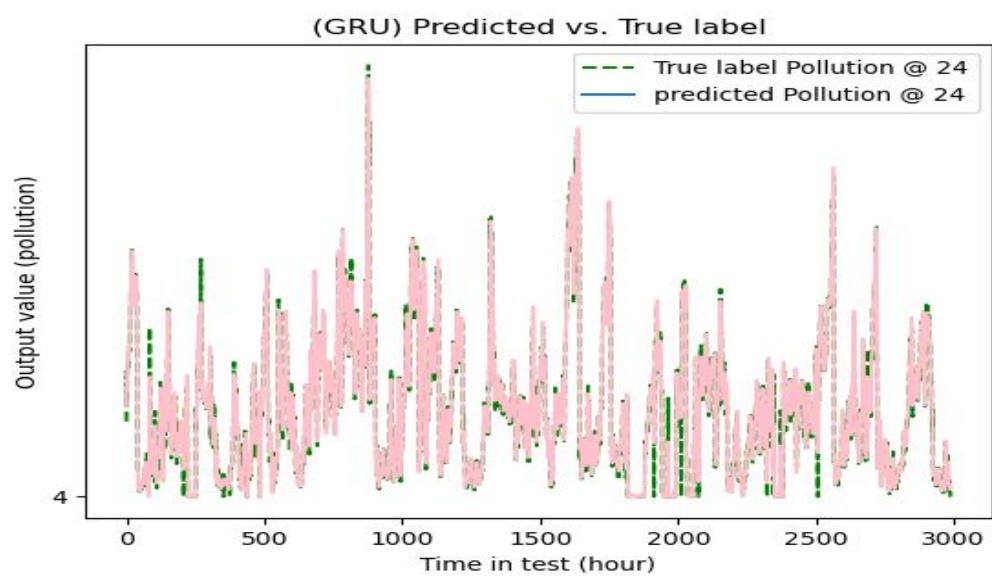
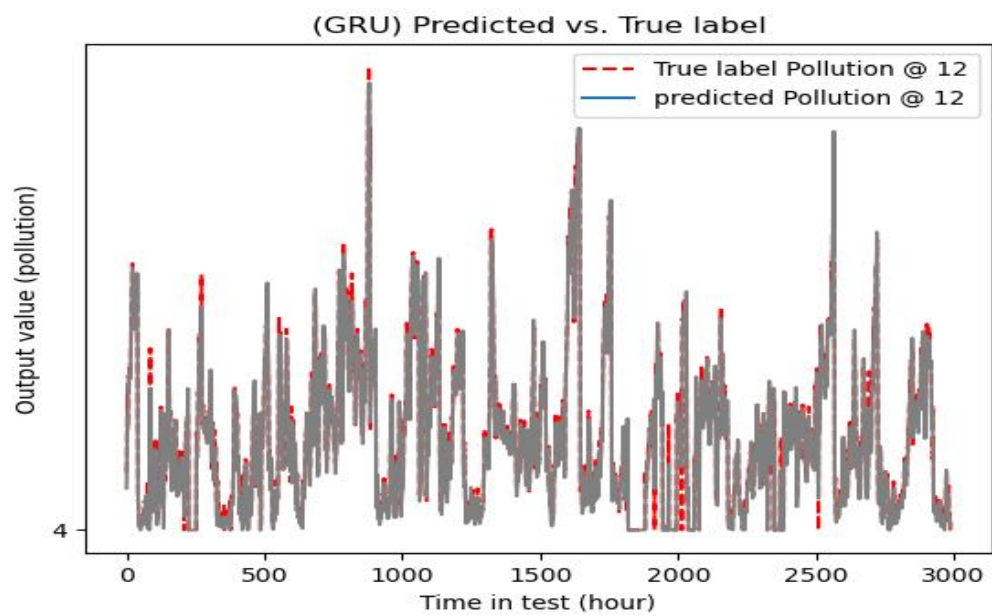
نمودار خطای داده ی ارزیابی برای 3 شبکه



نمودار مقدار حقیقی و پیش بینی برای شبکه rnn



نمودار مقدار حقیقی و پیش بینی برای شبکه lstm



نمودار مقدار حقیقی و پیش بینی برای شبکه gru

قسمت دوم (نقصان دادگان)

بخش 1 و 2 :

در این قسمت 12 هزار داده به عنوان داده train در نظر گرفتیم و به صورت جداگانه برای هر ستون (ویژگی) 20 درصد از داده ها را حذف کردیم.

بخش 3 :

1. Mean or Median

با محاسبه میانگین یا میانه اعداد برای مقادیر از دست رفته در یک ستون و سپس جایگزین کردن مقادیر از دست رفته در هر ستون بطور جداگانه این کار انجام می شود. این روش فقط با داده های عددی قابل استفاده است.

معایب این روش:

محاسبه میانگین یا میانه ارتباط بین ویژگی ها را در نظر نمی گیرد و فقط یک ویژگی را در نظر می گیرد و این روش به اندازه کافی دقیق نیست و ممکن است به داده ها بایاس وارد کند.

2. Interpolation

روشهای درون یابی (Interpolation) مجموعه ای از مدل های مختلف ریاضی و آماری را برای پیش بینی مقادیر نامعلوم بکار می برد. درون یابی برای داده های از دست رفته یک تابع را در نظر می گیرد تابعی که دقیقاً از یک مجموعه نقطه داده عبور می کند. این روش دقت بالایی دارد.

3. Drop value

در این روش تمام مقادیر ویژگی مورد نظر را صفر یا null می گذاریم و این ویژگی را برای پیش بینی کردن در نظر نمی گیرند که روش مناسبی نیست .

4. KNN

KNN الگوریتمی است که برای تطبیق یک نقطه با نزدیکترین همسایگان خود در یک فضای چند بعدی مفید است. این می تواند برای داده های پیوسته ، گسسته ، ordinal و طبقه ای استفاده شود که این روش برای بازبینی داده های از دست رفته مفید است.

در این روش داده از دست رفته بر اساس چند نقطه نزدیک به آن تخمین زده می شود.

فرض استفاده از KNN برای مقادیر از دست رفته این است که براساس متغیرهای دیگر ، می توان مقدار نقطه را توسط مقادیر نقاط نزدیک به آن تقسیم کرد.

پارامترهای knn :

1. تعداد همسایه ها

2. محاسبه فواصل بین نقاط

بخش 4 :

با سه روش mean, median و interpolate مقادیر حذف شده را پیش بینی نموده ایم. و در جدول شماره 1-2 نتایج مشخص شده است.

جدول 1-2 - محاسبه خطای mse

interpolate	median	mean	
0.00012488446121096606	0.0025670003274446233	0.0023236732469314077	pollution
3.851608517690169e-05	0.010500414421410288	0.011168809948515625	dew
4.5131519287054355e-05	0.010425221397068581	0.010310826496274652	temp
1.4357713340665556e-05	0.0072337236835031335	0.008069213377039098	pressure
0.016074331101566686	0.03490740809065327	0.03569888591147047	Wind-dir
9.150897694805283e-05	0.0026503526878885217	0.00219458968152935 1	Wind-spд
1.3282402765996284e-05	0.00044455876300357575	0.00044625557948163884	snow
7.877585995095972e-05	0.00048527521140578347	0.000492406112902226	rain

بخش 5:

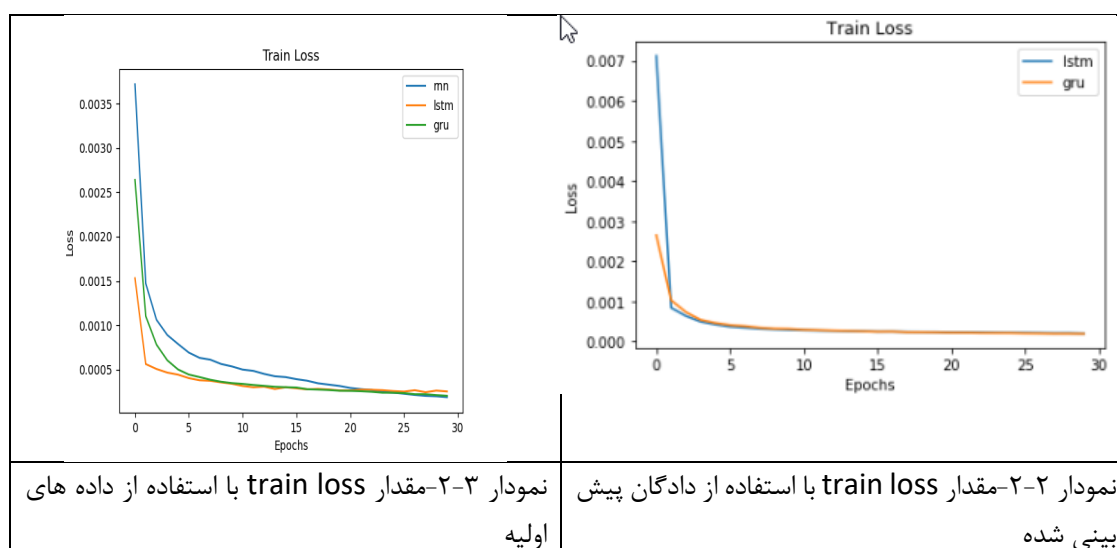
داده های از دست رفته برای هر ویژگی (ستون) به طور جداگانه و با سه روش میانه و میانگین و درون یابی پیش بینی شده اند و نتایج در جدول نشان داده شده است. پیش بینی داده ها با **روش Interpolation** مقدار خطای کمتری نسبت به میانگین و میانه دارد.

بخش 6

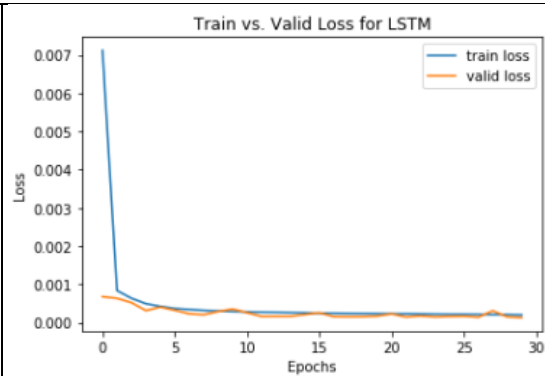
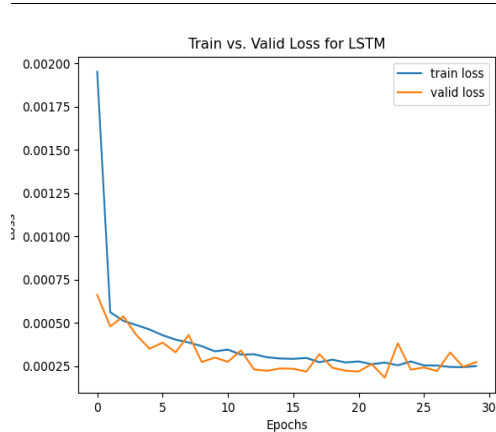
ستون سمت راست نمودارهای مربوط به داده های پیش بینی شده به روش Interpolation است و ستون سمت چپ نمودارهای مربوط به داده ای اولیه می باشد. همانطور که در شکل های زیر مشاهده می کنید اختلافی بین دقت داده های اولیه و داده های پیش بینی شده وجود ندارد و این بدین معنی است که داده ها به خوبی پیش بینی شده اند.

در نمودار 2-6 و 2-7 مقدار حقیقی و پیش بینی برای داده های پیش بینی شده برای میزان آلودگی روزانه با استفاده از شبکه lstm رسم شده است .

در نمودار 2-8 و 2-9 مقدار حقیقی و پیش بینی برای داده های پیش بینی شده برای میزان آلودگی روزانه با استفاده از شبکه gru نشان داده شده است و همانطور که از نمودار ها مشخص است شبکه قادر به پیش بینی خوبی بوده است.

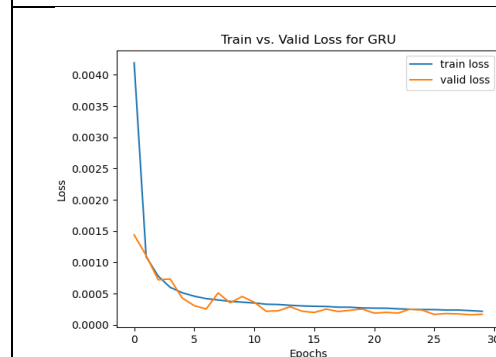


<p>نمودار ۲-۵- مقدار test loss با استفاده از داده های اولیه</p>	<p>نمودار ۲-۴- مقدار test loss با استفاده از دادگان پیش بینی شده</p>
<p>نمودار ۲-۷- نمودار مقدار حقیقی و پیش بینی برای داده های اولیه</p>	<p>نمودار ۲-۶- نمودار مقدار حقیقی و پیش بینی برای داده های پیش بینی شده</p>
<p>نمودار ۲-۹- نمودار مقدار حقیقی و پیش بینی برای داده های اولیه</p>	<p>نمودار ۲-۸- نمودار مقدار حقیقی و پیش بینی برای داده های پیش بینی شده</p>



نمودار ۱۱-۲- مقدار $loss$ شبکه $lstm$ برای داده های اولیه

نمودار ۱۰-۲- مقدار $loss$ شبکه $lstm$ برای داده های پیش بینی شده



نمودار ۱۳-۲- مقدار $loss$ شبکه gru برای داده های اولیه

نمودار ۱۲-۲- مقدار $loss$ شبکه gru برای داده های پیش بینی شده