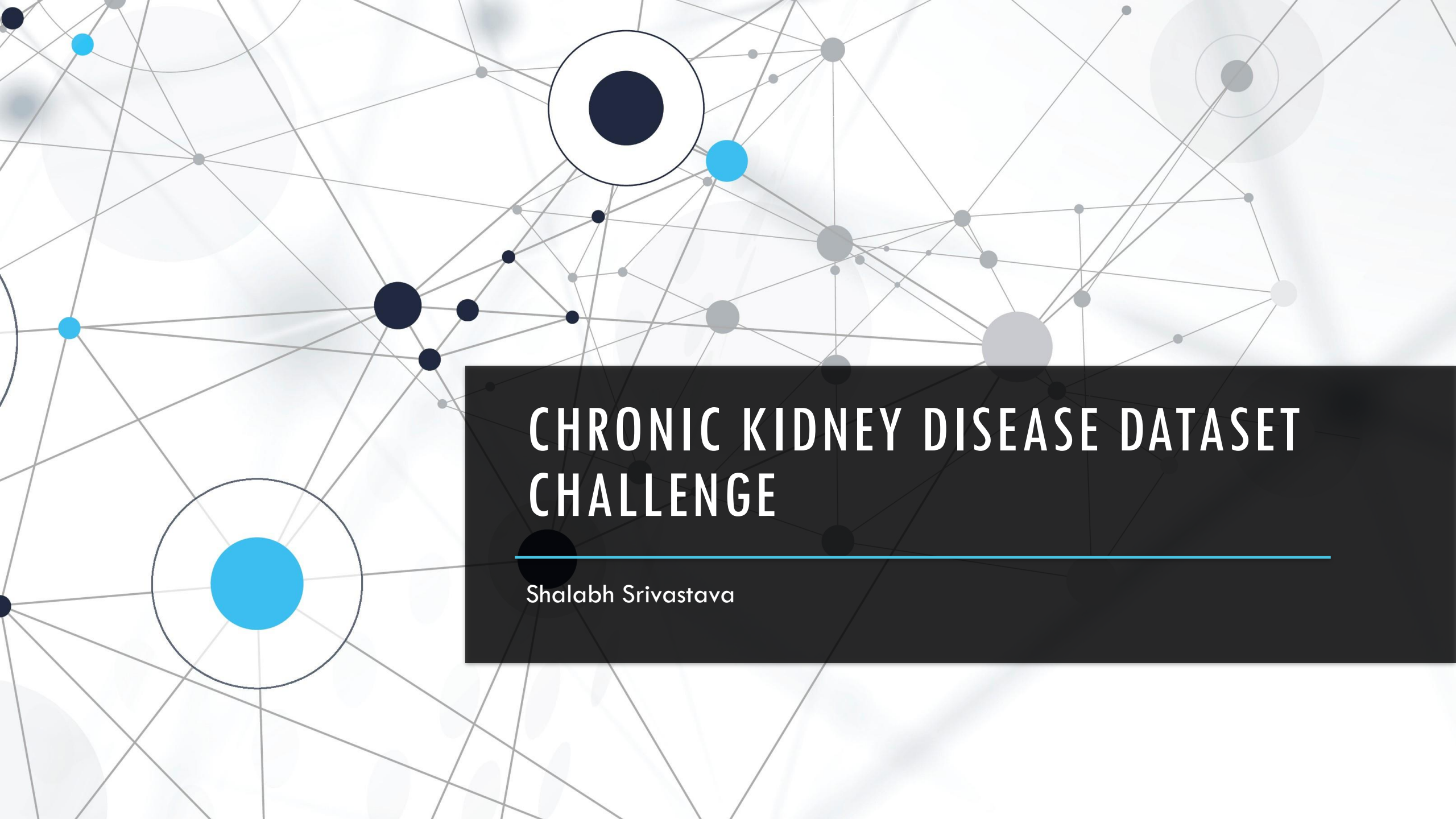


CHRONIC KIDNEY DISEASE DATASET CHALLENGE

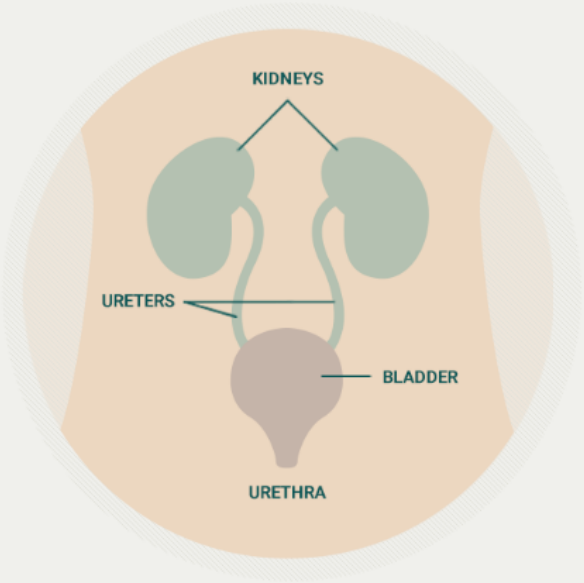
Shalabh Srivastava

OUTLINE

- About Kidneys
- Challenge overview
- Data pre-processing
- Model description implementation
- Result analysis
- Conclusion
- References

LEARN MORE ABOUT

WHAT KIDNEYS DO FOR YOUR BODY



- REMOVE WASTE & EXTRA WATER BY MAKING URINE
- HELP KEEP PHOSPHORUS & POTASSIUM LEVELS NORMAL
- KEEP BLOOD PRESSURE NORMAL
- KEEP BONES HEALTHY AND STRONG
- KEEP YOUR BODY CHEMICALS IN BALANCE
- CLEAN BLOOD & HELPS KEEP RED BLOOD CELL COUNTS NORMAL

ABOUT KIDNEYS

HOW KIDNEY WORKS?

- Kidneys: 2 located at either side of spine at the lowest level of rib cage.
- Building block of Kidney:
  - Nephron (about 1 million in each kidney), decreases with age or disease.
- A nephron consists of a filtering unit of tiny blood vessels called **a glomerulus** attached to **a tubule.**

CHRONIC KIDNEY DISEASE (CKD)

Chronic kidney disease is defined as having some type of kidney abnormality, which decreases Kidney's function.

Some causes of CKD:

- Other diseases such as Diabetes, High Blood Pressure, etc. (most common type)
- Hereditary
- Congenital
- Use of drugs and toxins

DETECTION OF CKD

Early detection and treatment of chronic kidney disease are the keys to keeping kidney disease from progressing to kidney failure.

Laboratory Tests:

- Albumin to Creatinine Ratio (ACR), estimates the amount of albumin that is in your urine.
- Test for blood Ceratinine
- Blood Pressure
- estimated glomerular filtration rate

Many forms of kidney disease do not produce symptoms until late in the course of the disease.

TO PREDICT WHETHER A PATIENT WILL PROGRESS IN CKD STAGING BASED ON THE PATIENT'S PAST ELECTRONIC HEALTH RECORD INFORMATION.

DATA AVAILABLE

Dataset available for 300 patients

Personal data for each patient –

- Stage progress
- Gender
- Race
- Age

Time dependent electronic health record (EHR) information –

- Creatinine
- Diastolic blood pressure
- Systolic blood pressure
- Glucose
- Haemoglobin
- LDL cholesterol

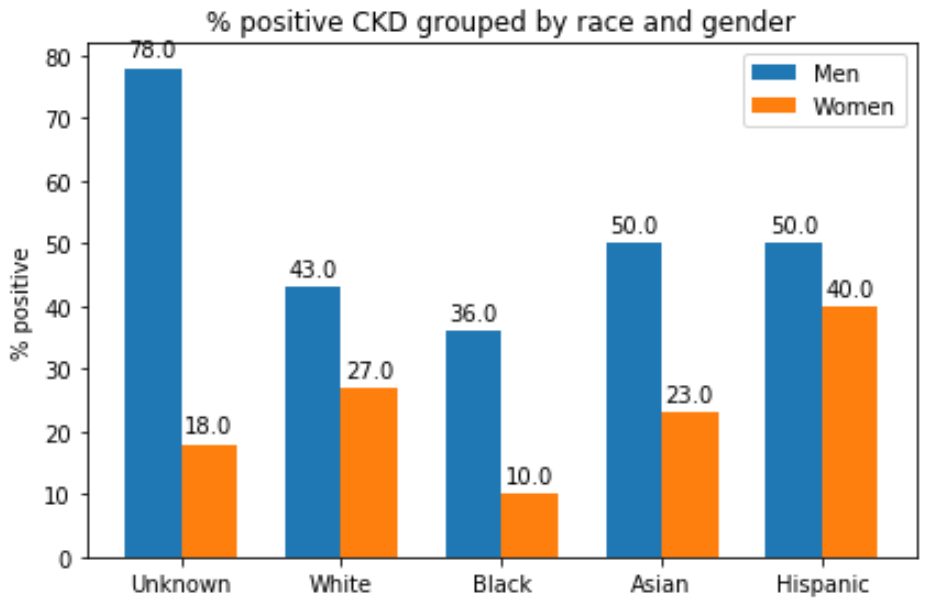
Time period of medicine administered by each patient

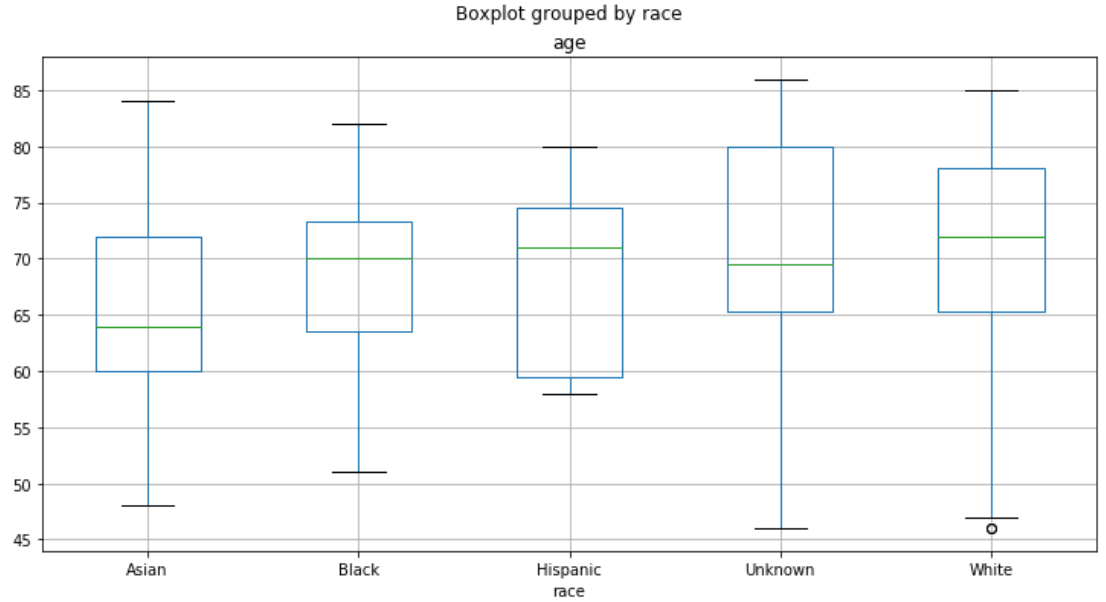
PATIENT'S PROFILE

CKD strongly depends on race, sex and age

Men and patient's in the age bracket of 60-80 are more prone to CKD

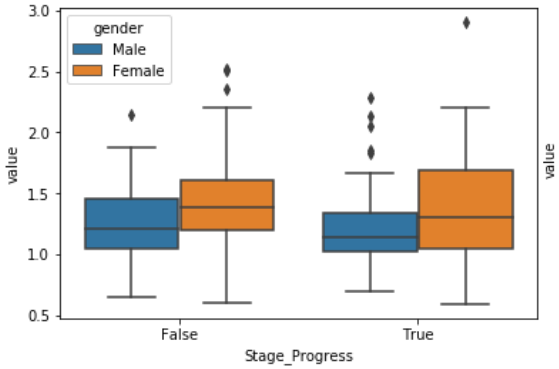
Important to add patient's profile to time series classification

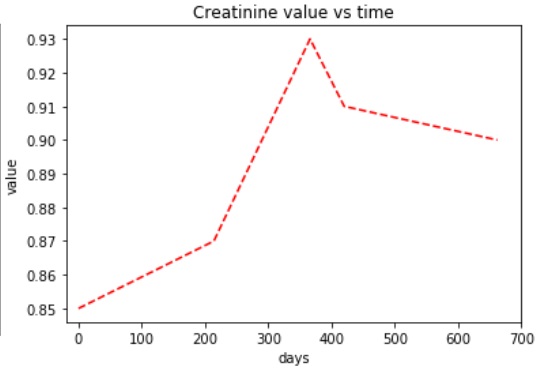




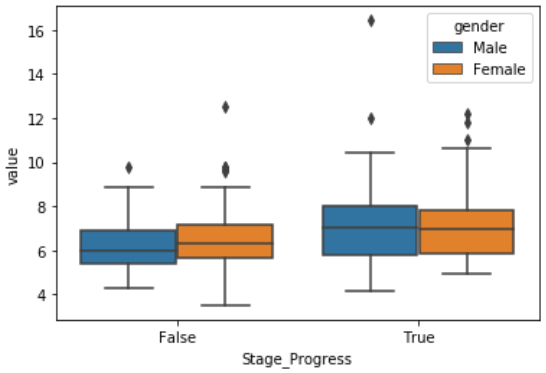
EHR

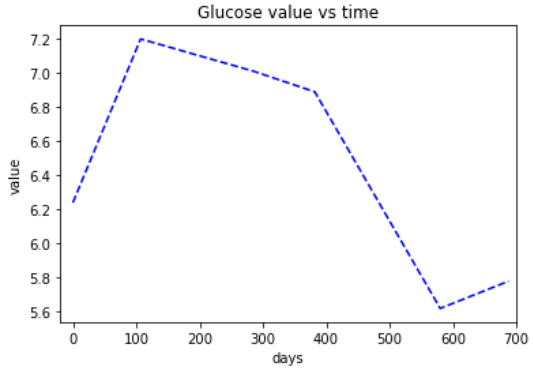
Creatinine	A chemical waste product of normal muscle function, in the blood that passes through the kidneys to be filtered and eliminated in urine.
Dependent on	• Gender • Activity Level • Body Size • Medication
Relation to CKD	Due to CKD, Creatinine level rises in blood while decreases in urine
Tests	Blood: Serum Creatinine Urine: Creatinine clearance
Normal Range	0.8 to 1.4 milligrams per deciliter (mg/dl).





Glucose	When blood sugar levels get too high, the condition is called hyperglycemia. If it gets too low, it is called hypoglycemia.
Dependent on	• Eating habits • Physical activity • Medications
Relation to CKD	High blood sugar from diabetes can cause damage inside kidneys CKD and diabetes together, increase the risk of low blood sugar
Test	Multiple ways to test (home kits) A1c — Glycosylated hemoglobin test used for CKD
Normal Range	3.9 to 7.1 mmol/L (70 to 130 mg/dL for non diabatic (with fasting)

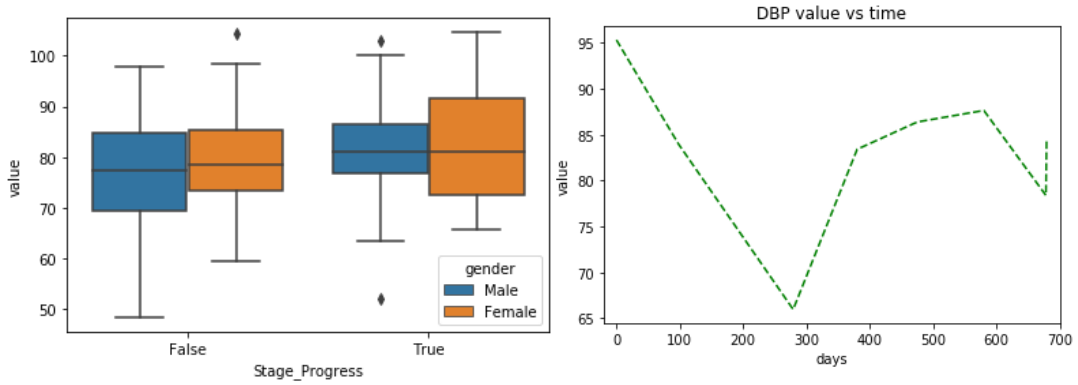


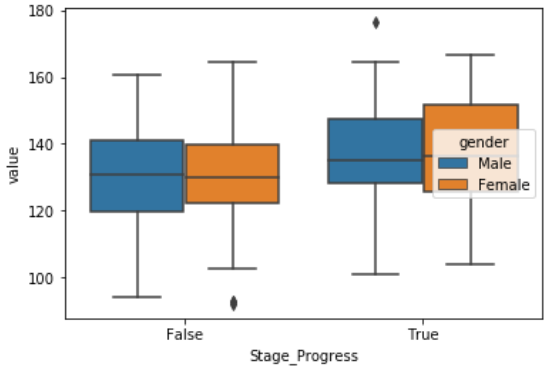


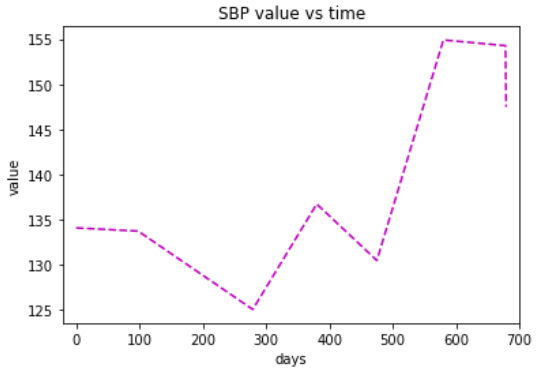
EHR

SBP	Systolic blood pressure measures the pressure that blood exerts on vessels while the heart is beating.
DBP	Diastolic blood pressure measures the pressure between heartbeats.

Dependent on	• Renin (hormones)	• Stress
	• Exercise	• Eating habits
Relation to CKD	High blood pressure (BP) can lead to slow damage in the blood vessels and capillary in the Kidney Once a person has CKD, there will be increased renin in blood, which increases BP	
Test	A blood Pressure cuff around patient's arm	
Normal Range	The optimal systolic blood pressure is 120 mmHg. The optimal diastolic blood pressure is 80 mmHg.	

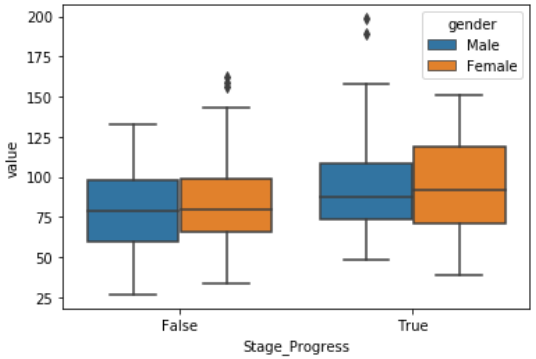


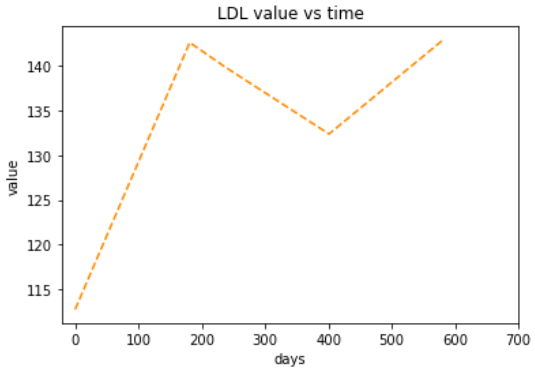




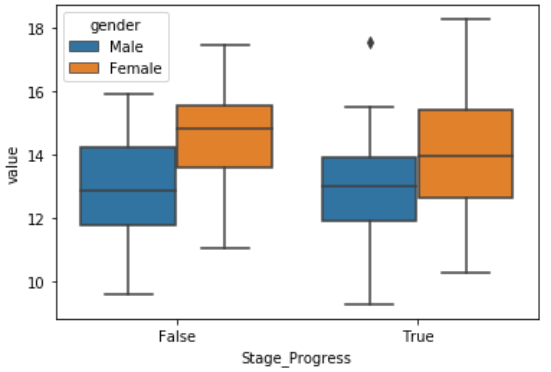
EHR

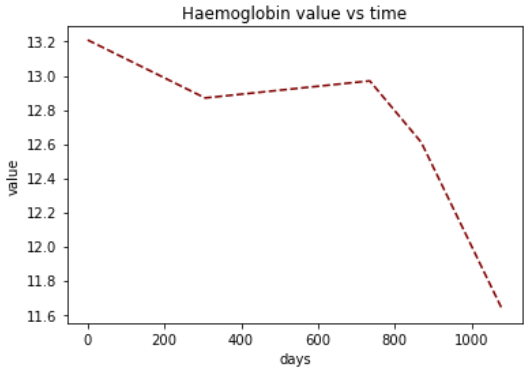
Low-density lipoprotein	LDL cholesterol is often called the “bad” cholesterol because it collects in the walls of your blood vessels.
Dependent on	• Amount of LDL produced by body, impacted by diet, medication • Rate at which LDL is sent back to liver to be destroyed
Relation to CKD	LDL cholesterol plaque can also clog the renal arteries and cut off blood flow to the kidneys, resulting in loss of kidney function.
Test	Testing for LDL-C involves calculating of LDL-C amount in blood based on results of a lipid panel.
Normal Range	Less than 100 milligrams per deciliter (mg/dL): Optimal



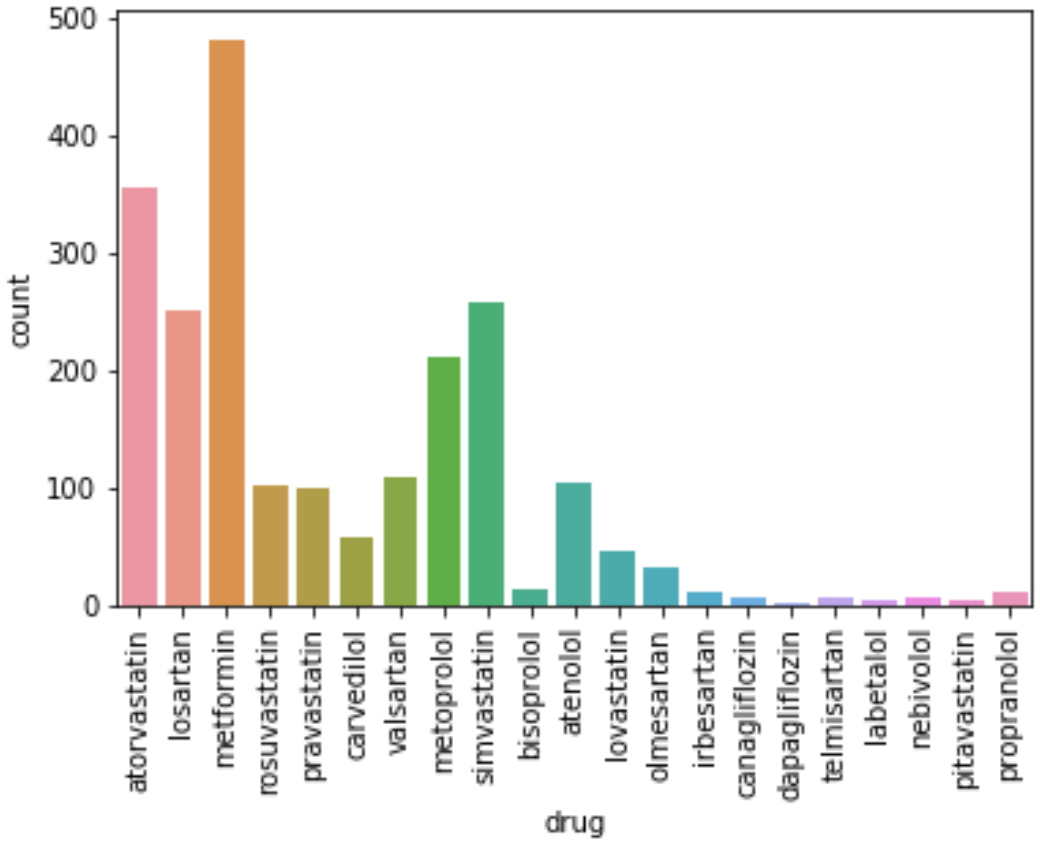


Haemoglobin	Anaemia (low haemoglobin) is a condition in which the body has fewer red blood cells than normal.
Dependent on	• Gender • Age • Altitude • Smoking Habits
Relation to CKD	CKD cause decrease in erythropoietin (EPO) by the Kidney EPO prompts the bone marrow to make red blood cells The bone marrow makes fewer red blood cells, causing anaemia
Test	Routine blood count
Normal Range	In an adult, 12 to 18 grams per deciliter of blood.





MEDICINE



Cholesterol:

Atorvastatin, Rosuvastatin, Pravastatin, Simvastatin, Lovastatin, Pitavastatin

Blood Sugar:

Metformin, Canagliflozin, Dapagliflozin

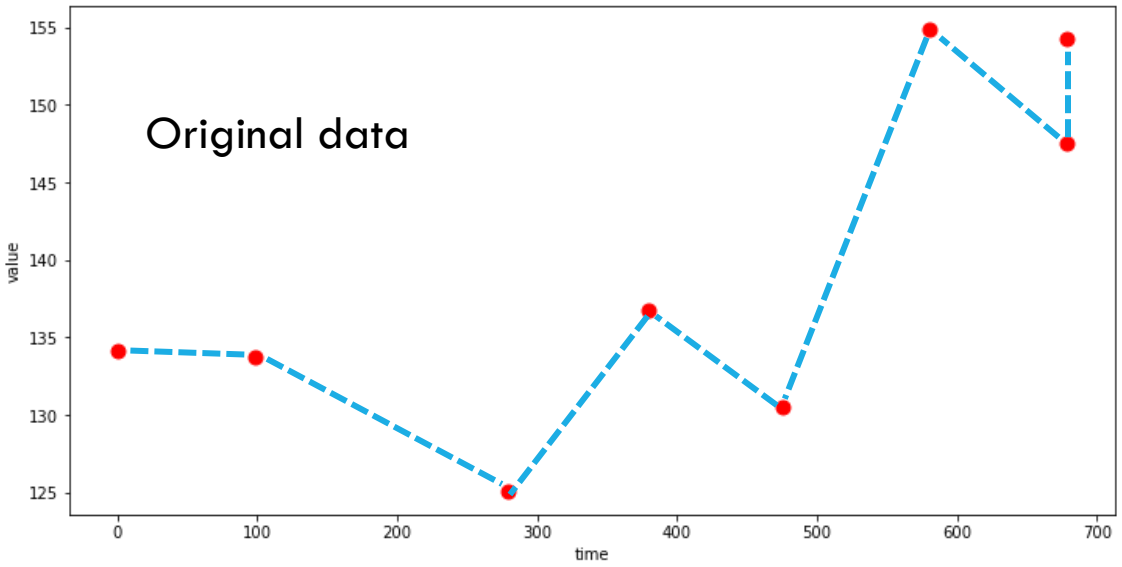
Blood Pressure:

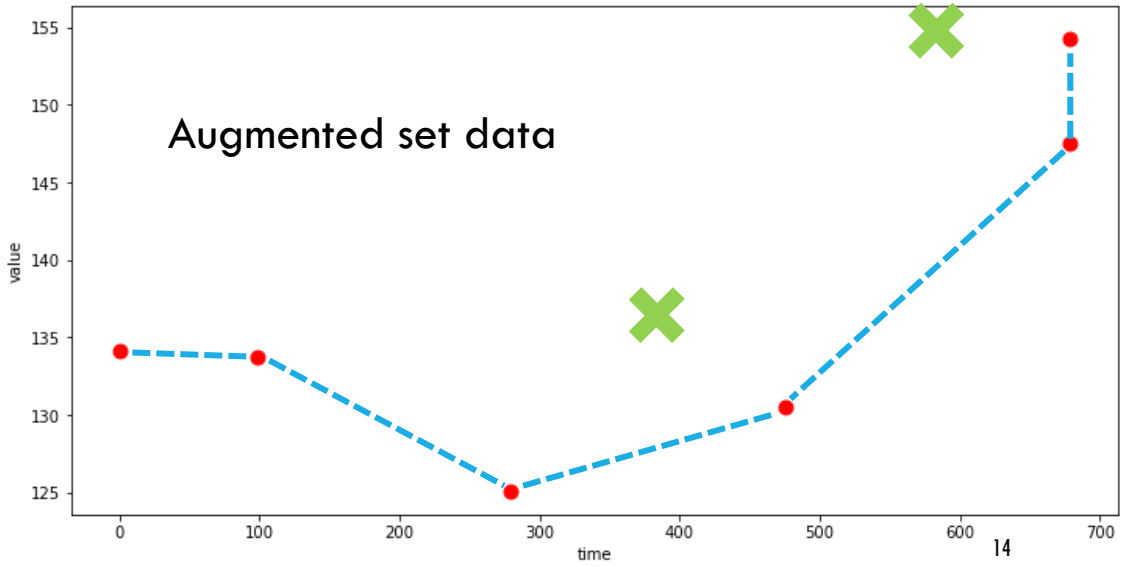
Losartan, Carvedilol, Valsartan, Metoprolol, Bisoprolol, Atenolol, Olmesartan, Irbesartan, Telmisartan, Labetalol, Nebivolol, Propranolol

DATA PRE-PROCESSING

- Dividing dataset into train, valuation and test dataset

- Removing $1/4^{th}$ of EHR data points at to produce augmented dataset
- Linearly interpolating data across 700 days





DATA PRE-PROCESSING

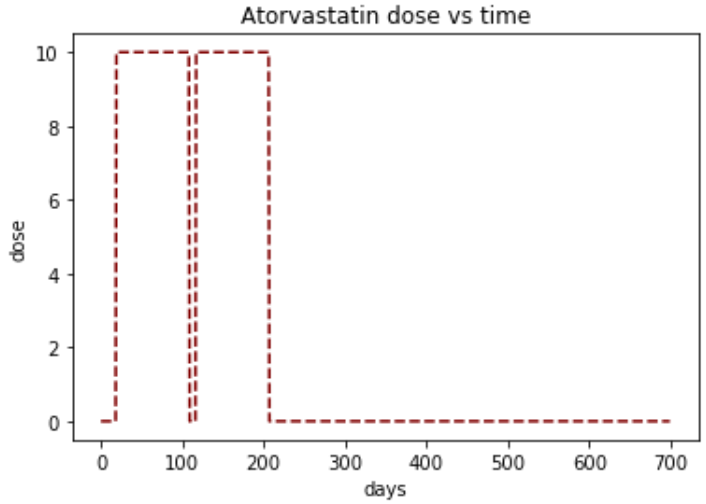
Creating one hot vectors for race and gender

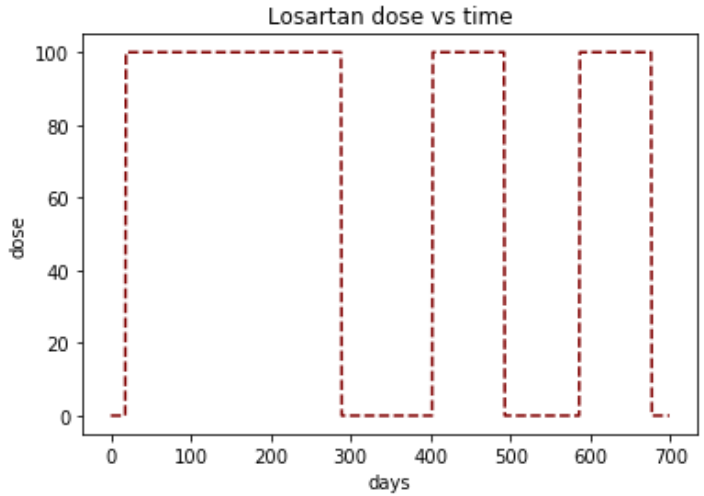
Vectorization and normalization of age

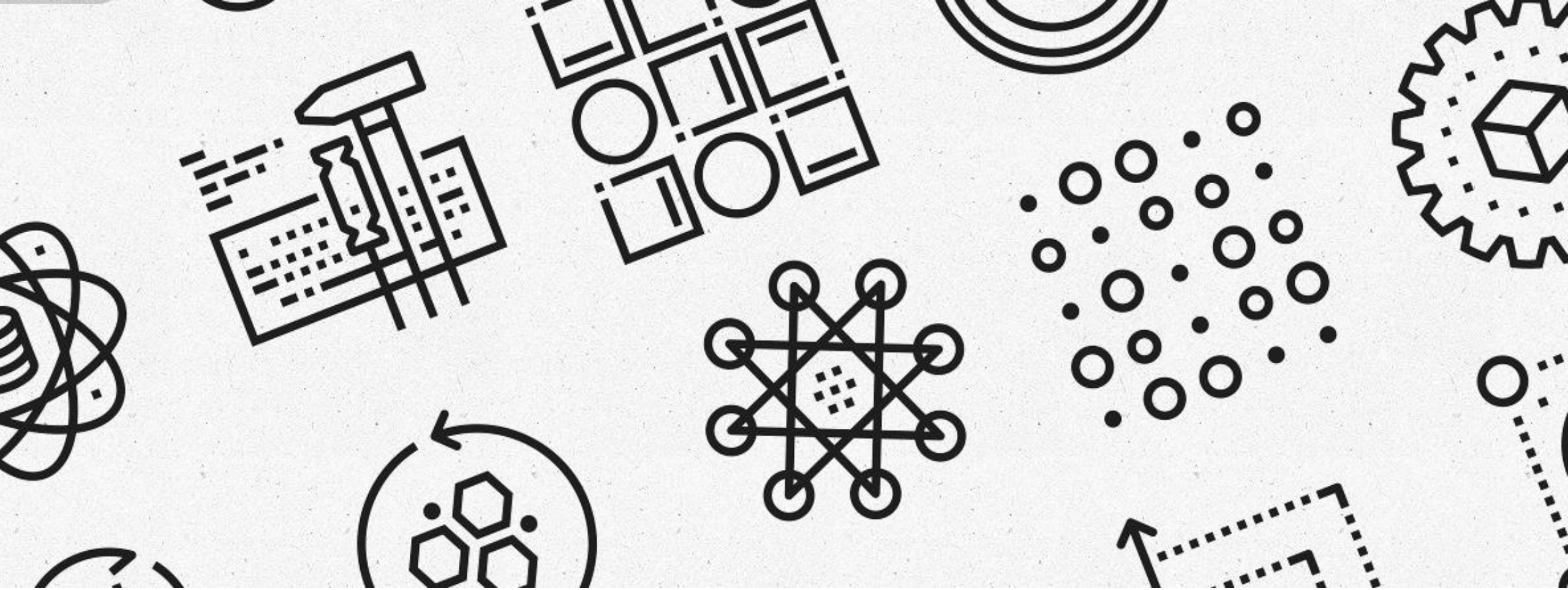
```
for num, race in enumerate(['Unknown', 'White', 'Black', 'Asian', 'Hispanic']):
    if (T_demo[(T_demo['id']==idx1)]['race']==race).bool():
        Demo[idx1,num]=1
if (T_demo[(T_demo['id']==idx1)]['gender']=='Male').bool():
    Demo[idx1,5] = 1
Demo[idx1,6] = (T_demo[(T_demo['id']==idx1)]['age']-min_age)/del_age
```

Vectorization of medicine dosage

MinMaxScaler to normalize the data



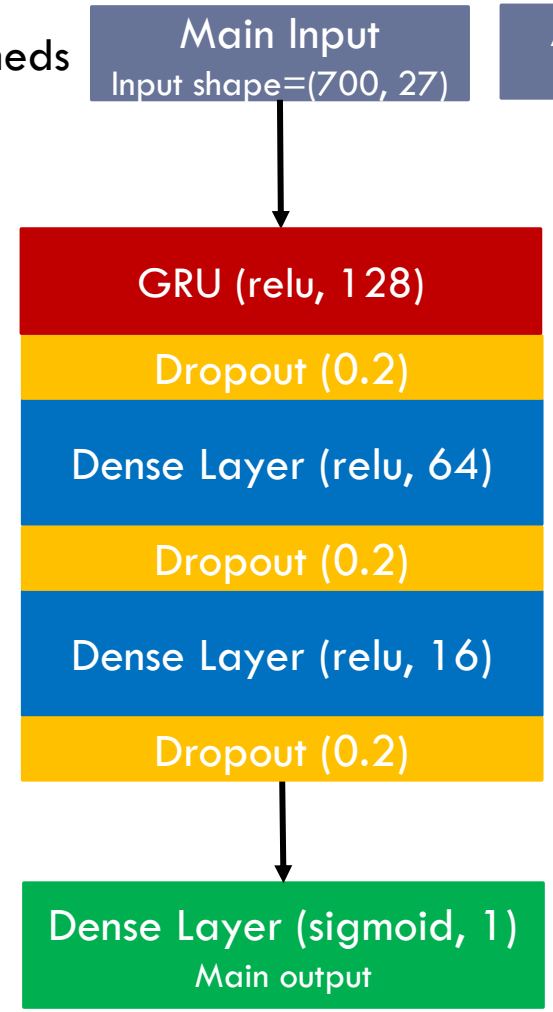




MODEL

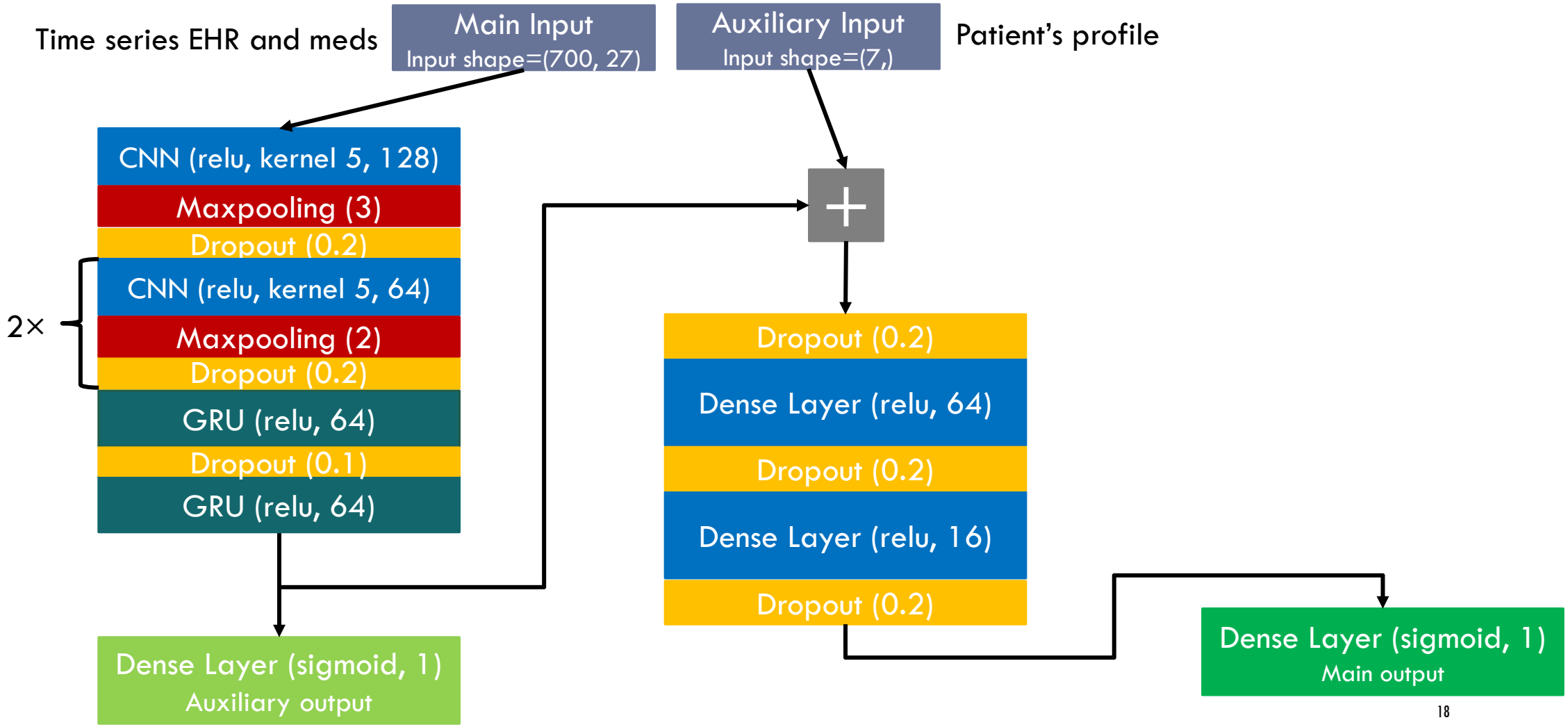
MODEL1: GRU+DENSE

Time series EHR and meds

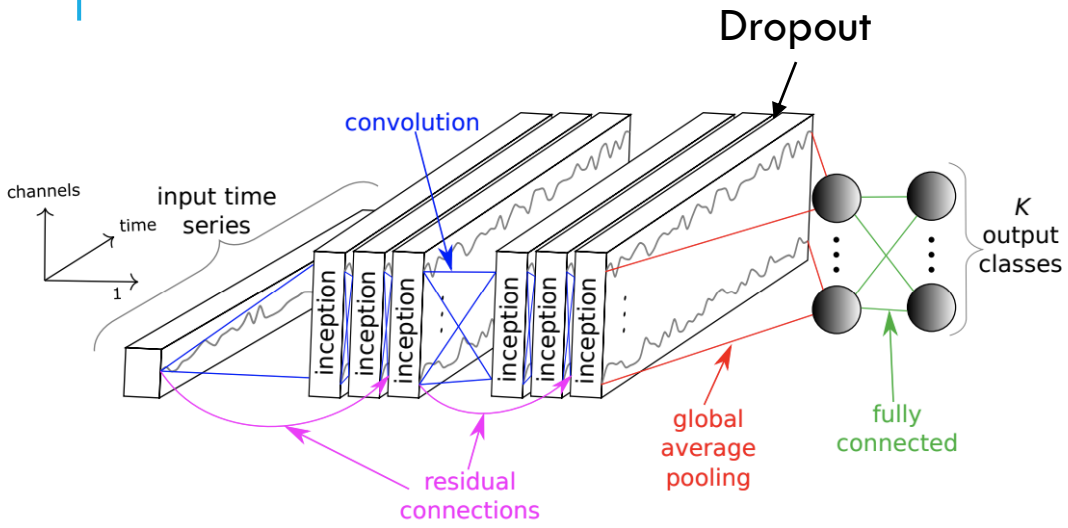


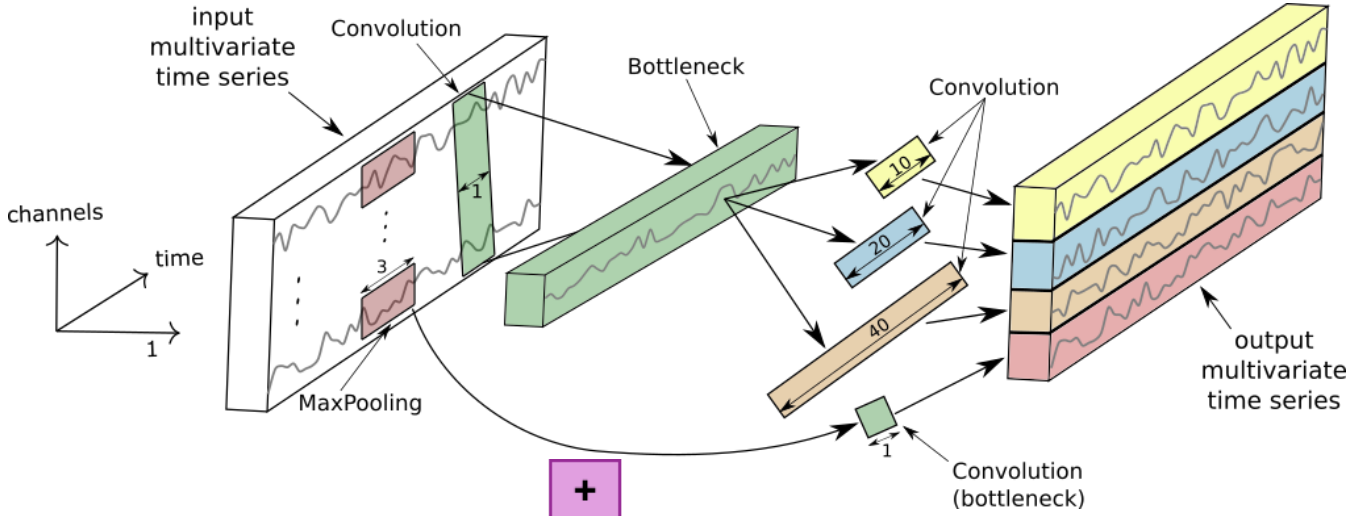
Patient's profile

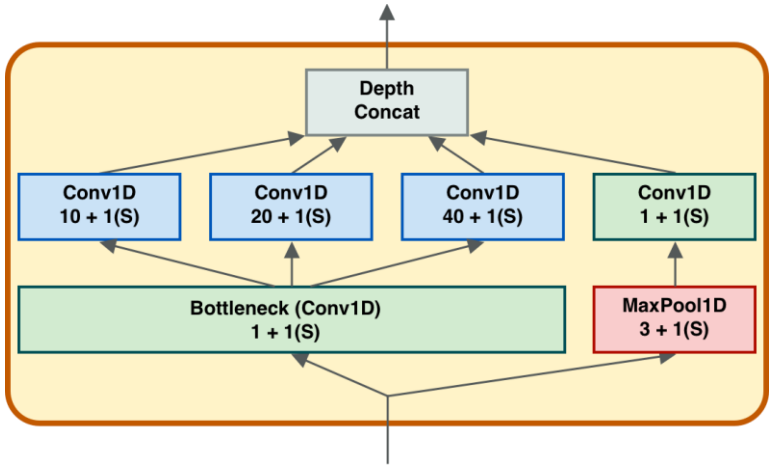
MODEL2: CNN+GRU+DENSE

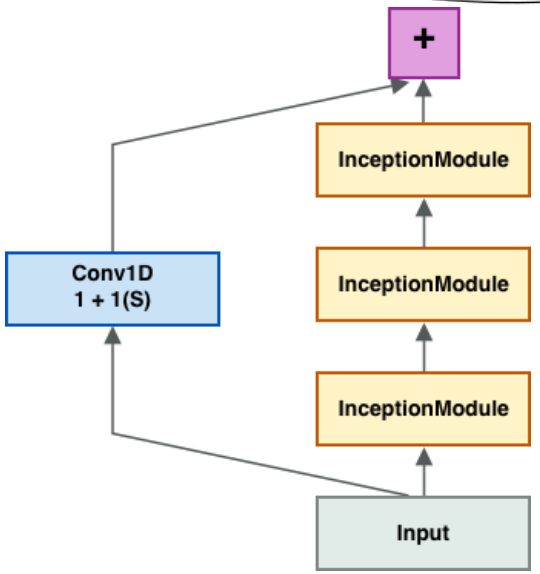


MODEL3: INCEPTIONTIME



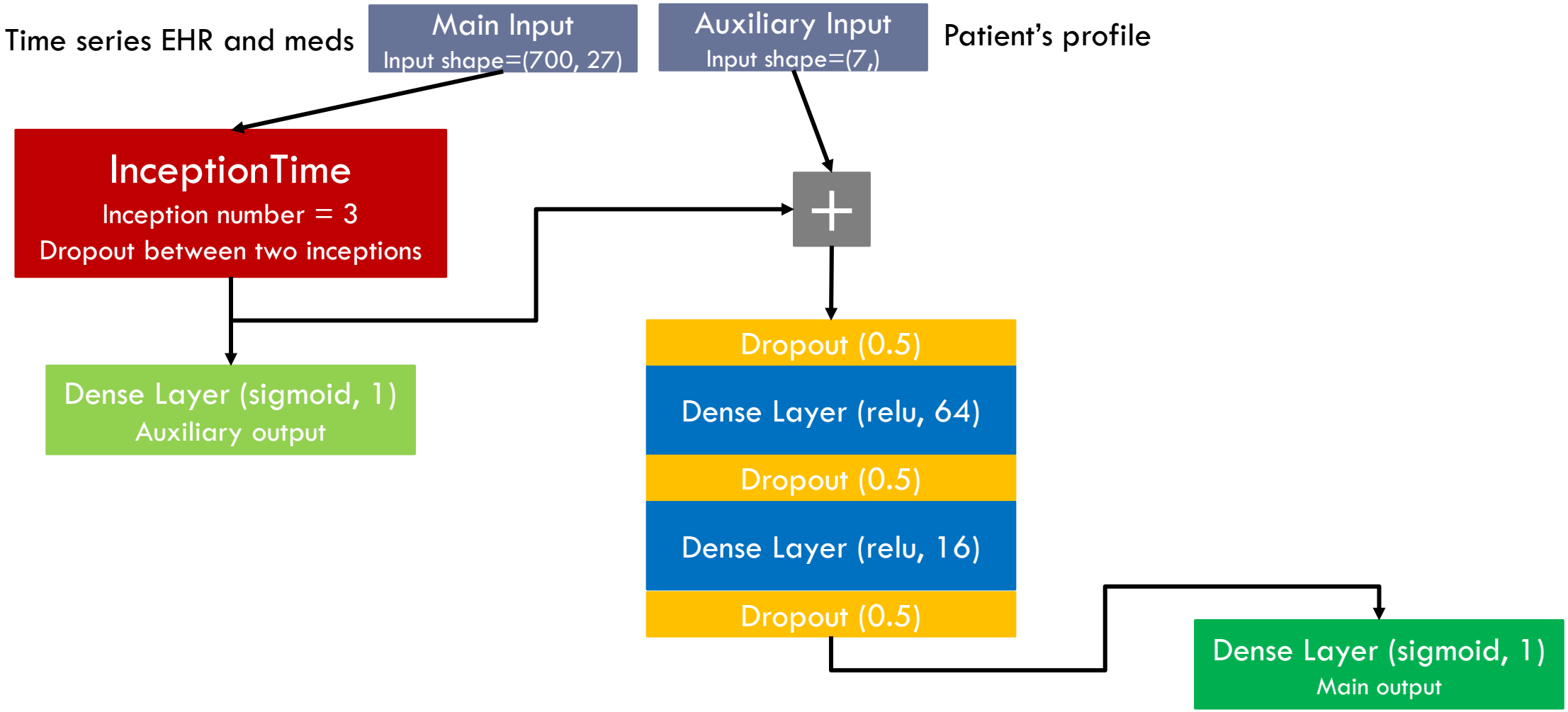






InceptionTime: Finding AlexNet for Time Series Classification, Fawaz *et.al.*, arXiv:1909.04939 (2019)

MODEL3: INCEPTIONTIME (MODIFIED)



MODEL COMPILATION

MODEL:1

```
Total params: 69,601
Trainable params: 69,601
Non-trainable params: 0
```

MODEL:2

```
Total params: 134,561
Trainable params: 134,561
Non-trainable params: 0
```

MODEL:3

```
Total params: 248,897
Trainable params: 247,873
Non-trainable params: 1,024
```

```
model.compile(loss='binary_crossentropy', optimizer=tf.keras.optimizers.Adam(), metrics=['accuracy', f1_factor])

reduce_lr = tf.keras.callbacks.ReduceLROnPlateau(monitor='loss', factor=0.5, patience=50, min_lr=0.001)

model_checkpoint = tf.keras.callbacks.ModelCheckpoint('model', monitor='loss', restore_best_weights=True)

callbacks = [reduce_lr, model_checkpoint]
```

Optimizer: Adam + ReduceLROnPlateau (callback)

ReduceLROnPlateau monitors 'loss' over a 'patience' number of epochs by reducing learning rate by a 'factor'

ModelCheckpoint saves checkpoints with the epoch number and the validation loss in the filename.

LOSS FUNCTION AND METRICS

Loss function used:

Binary Cross Entropy $= -(ylog(p) - (1-y)log(1-p))$

y is true label and p is predicted label

Metrics used:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN}$$

$$\text{F1 score} = \frac{2\times Prec\times Recall}{Prec+Recall}$$

$$\text{Prec} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Confusion Matrix

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

RESULTS AND ANALYSIS

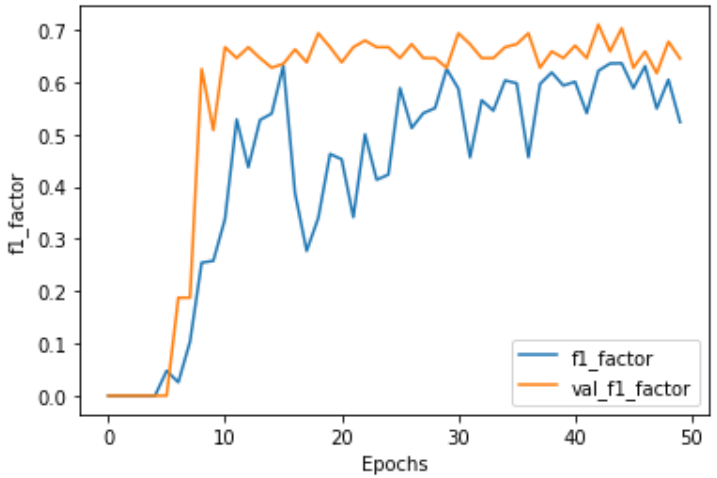
RESULTS AND ANALYSIS

Without Data Augmentation

MODEL:1

	loss	accuracy	F1
Train	0.5887	0.7500	0.6045
Valid.	0.6027	0.7167	0.6771

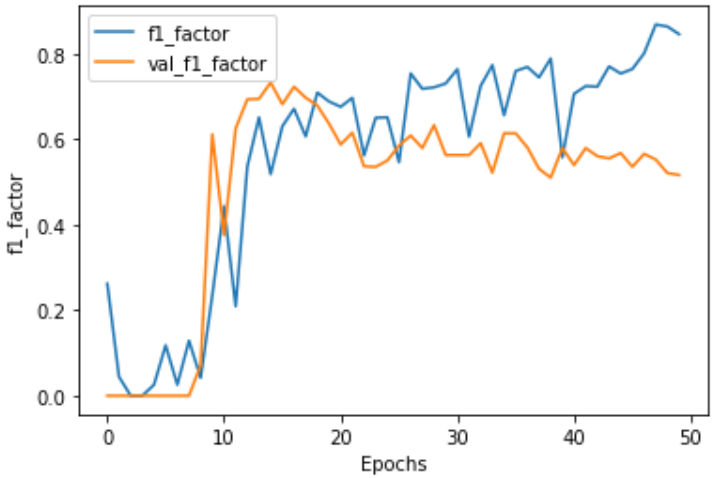
CKD Staging	Predicted Yes	Predicted No
Actual Yes	10	6
Actual No	13	31



MODEL:2

	loss	accuracy	F1
Train	0.9436	0.7444	0.6718
Valid.	0.8853	0.8000	0.7233

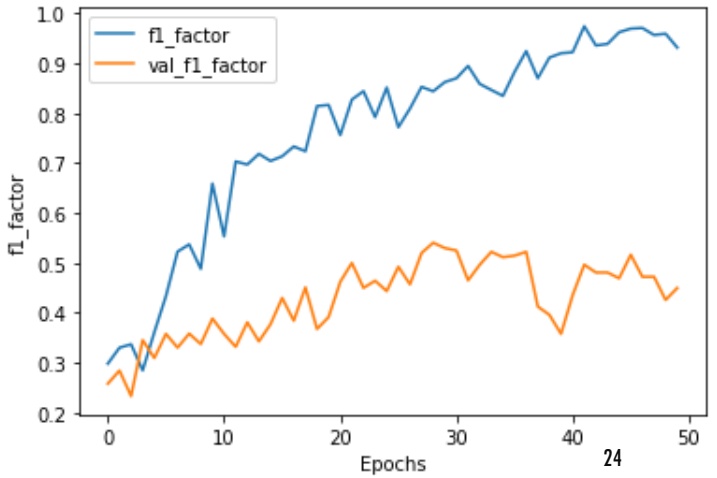
CKD Staging	Predicted Yes	Predicted No
Actual Yes	10	6
Actual No	21	23



MODEL:3

	loss	accuracy	F1
Train	0.3351	0.9111	0.8624
Valid.	0.7325	0.6500	0.5298

CKD Staging	Predicted Yes	Predicted No
Actual Yes	12	4
Actual No	20	24



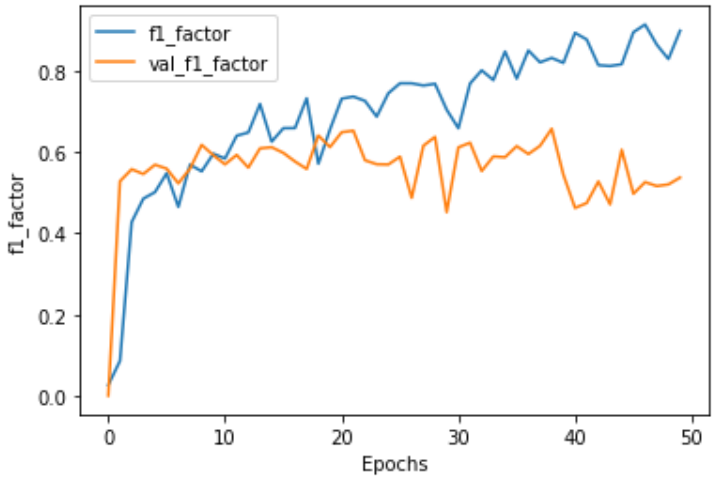
RESULTS AND ANALYSIS

With Data Augmentation

MODEL:1

	loss	accuracy	F1
Train	0.4744	0.7854	0.5906
Valid.	0.7143	0.6838	0.5689

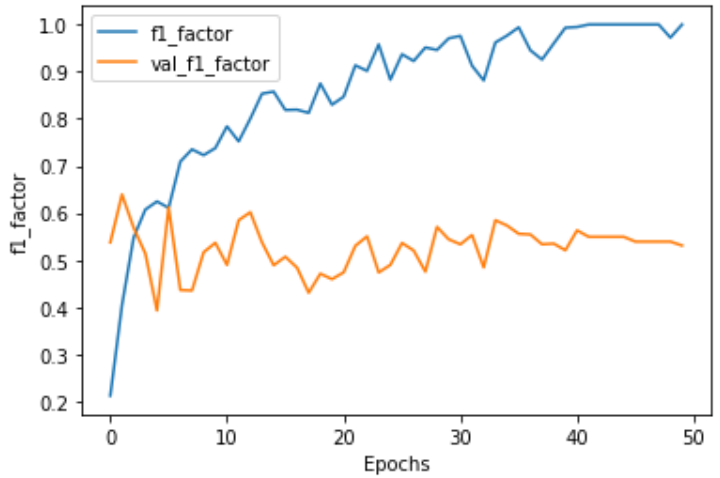
CKD Staging	Predicted Yes	Predicted No
Actual Yes	47	51
Actual No	80	198



MODEL:2

	loss	accuracy	F1
Train	0.7153	0.7996	0.6651
Valid.	2.1166	0.6812	0.6052

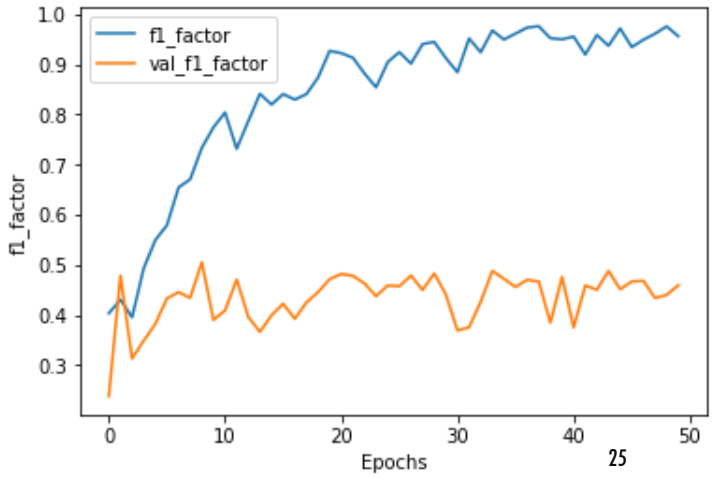
CKD Staging	Predicted Yes	Predicted No
Actual Yes	73	25
Actual No	109	169



MODEL:3

	loss	accuracy	F1
Train	0.3717	0.8362	0.7109
Valid.	0.7550	0.6812	0.4917

CKD Staging	Predicted Yes	Predicted No
Actual Yes	52	46
Actual No	109	169



CONCLUSION

- Model 1 shows best per performance for without AD on test dataset (F1 = 0.512)
- Model 2 shows best per performance for with AD on test dataset (F1 = 0.521)
- Data Augmentation degrades the performance of models on validation sets
- Model 3 shows large over fitting

Scope of improvement:

- Increase in number of datapoints can improve the performance of the models
- Results have shown overfitting – optimising of regularizing parameters
- Improve the method for data augmentation

REFERENCES

https://kidneyfailurerisk.com/

www.kidney.org

https://www.ahajournals.org/doi/10.1161/HYPERTENSIONAHA.119.13422

https://www.diabetes.co.uk/blood-glucose/what-affects-blood-glucose-levels.html