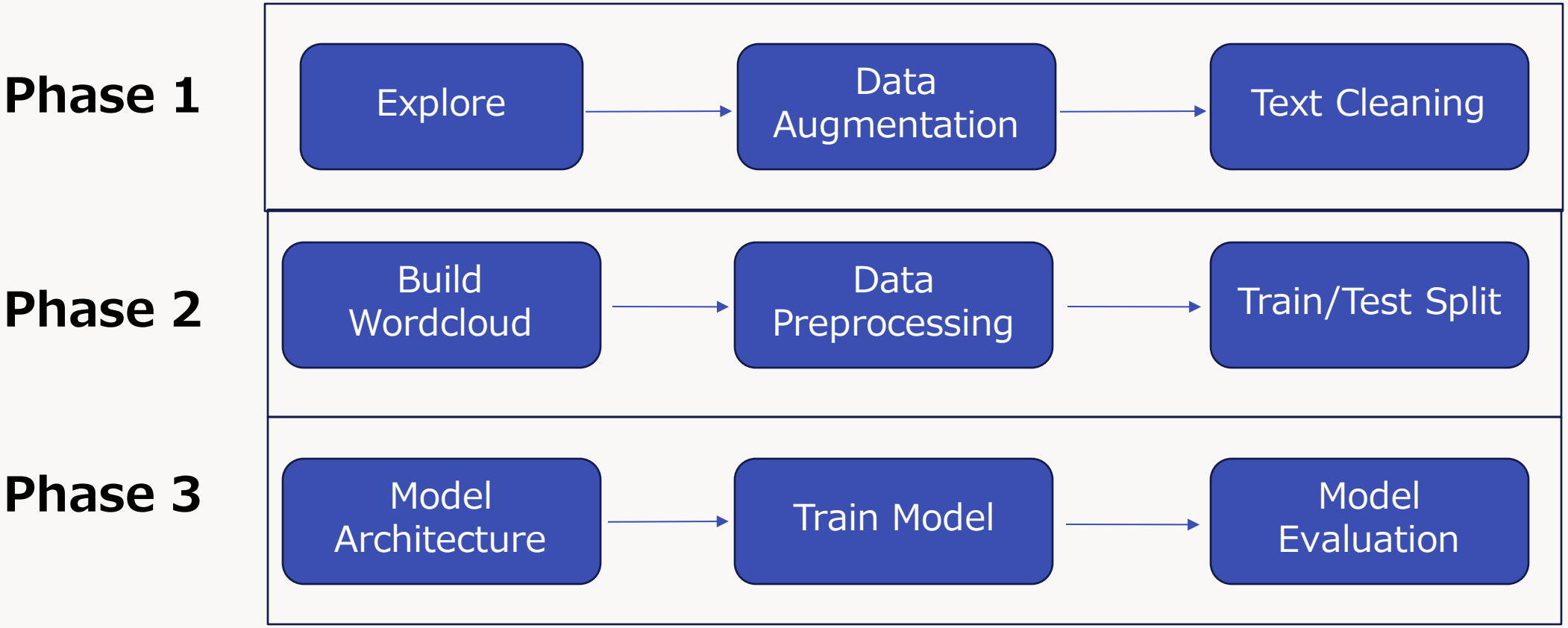
AAA RADIOLOGY REPORTS TO DETECT PRIOR SURGERY

detect if the patient had a prior surgery or not based on the radiology report

Abdelrahman Mohamed

Work Flow


Phase 1
Explore
Data Augmentation
Text Cleaning
Phase 2
Build Wordcloud
Data Preprocessing
Train/Test Split
Phase 3
Model Architecture
Train Model
Model Evaluation

Explore

Our Columns Information

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1514 entries, 0 to 1513
Data columns (total 5 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Unnamed: 0                    1514 non-null   int64
 1   PERSON_ID                     1514 non-null   object
 2   PROCEDURE_OCCURRENCE_ID_SCAN  1514 non-null   object
 3   Text                          1514 non-null   object
 4   Prior Surgery                 1514 non-null   bool
dtypes: bool(1), int64(1), object(3)
memory usage: 48.9+ KB
```

Summary

1- We should consider the "Text" column as our feature that should be fed to the model to predict the label that we assume it as "Prior surgery" column

2- we have only 72 True label so we have a huge bias in the labels so to solve this problem we can take a sample from False labels that close to True labels length and then use data augmentation methods to increase dataset samples

Labels insights

```
False_values ,True_values =df['Prior Surgery'].value_counts()
print(f"There is {False_values} False Value, and {True_values} True Value")
```

```
There is 1442 False Value, and 72 True Value
```

Abdelrahman Mohamed

Data Augmentation

Code Block

```
# Show difference between normal and augmented text
normal_words = []
aug_words = []
for i in range(len(df_new.clean_text.iloc[0].split(' '))):
    norm = df_new.clean_text.iloc[0].split(' ')[i]
    aug = df_new.clean_text.iloc[172].split(' ')[i]
    if aug != norm:
        normal_words.append(norm)
        aug_words.append(aug)
```

```
print(f"Normal Words: {normal_words}")
print(f"Augmented Words: {aug_words}")
```

```
Normal Words: ['lack', 'linear', 'trace', 'aneurysm', 'replacement']
Augmented Words: ['miss', 'elongate', 'suggestion', 'aneurism', 'renewal']
```

Augmentation usually add more words or replace exact words in the text without affecting the meaning so we use it to increase the samples we have in our dataset to avoid overfitting and generalize the prediction.

As we see in this example the lack replaced with miss and replacement with renewal etc.

Text Cleaning

1- removing stopwords
2- removing special characters
3- stemming the text
4- convert to lowercase
5- remove white spaces.

```
df.Text[0]
```

```
'\r\nEXAMINATION: CTA aortic dissection - chest, abdomen, and pelvis including iliac arteries\r\n\r\nIMPRESSION:  \r\n\r\nNew moderate left a
nd small right simple pleural effusions with associated bibasilar atelectasis. Otherwise, no significant change in the Stanford type B aortic
intramural hematoma compared to October 14, 2022.  \r\n\r\nCLINICAL INDICATION: 65 yo with AAA :: AAA (abdominal aortic aneurysm)\r\n\r\nTECH
NIQUE: \r\nCT scanning of the chest was performed without administration of intravenous contrast. Following this, CTA of the chest, abdomen a
nd pelvis was performed following intravenous administration of contrast. Reformats were constructed in the sagittal and coronal planes. MIP
images were performed and sent to PACS for archiving.  \r\nIntravenous contrast: IOPAMIDOL 76 % INTRAVENOUS SOLUTION: 100 mL \r\n\r\nCOMPARIS
ON: CT aortic dissection dated October 14, 2022.\r\n\r\nINTERPRETATION: \r\n\r\nSCOUT: No additional findings.\r\n \r\nAIRWAYS, LUNGS AND PLE
URA: The central tracheobronchial tree is patent. There are new moderate left and small right simple pleural effusions with associated bibasi
lar atelectasis. There is no pneumothorax. \r\n\r\nMEDIASTINUM: There are no enlarged mediastinal, hilar or axillary lymph nodes. The visuali
zed portion of the thyroid gland is unremarkable.    \r\n\r\nHEART AND VESSELS: The heart is normal in size. There are coronary artery and ao
rtic calcifications. There is left aortic arch with bovine variant branching.. The central pulmonary arteries have normal caliber. Again seen
is intramural hematoma extending from distal to the left subclavian artery through the level of the diaphragm. No contrast opacification of t
he false lumen is seen on arterial phase.. No change in size of the distal aortic arch/proximal descending aorta measuring up to 3.8 cm. Ther
e is no pericardial effusion. \r\n\r\nLIVER: Again seen is a 2 cm left lobe hepatic cyst, stable compared to prior study. \r\nBILIARY SYSTEM:
: Within normal limits. \r\nSPLEEN: Within n
structing stones. \r\n\r\nBOWEL/MESENTERY: N
itis.  The appendix is not seen; however, th
r\nREPRODUCTIVE ORGANS: No pelvic masses.\r\
ominal or pelvic lymphadenopathy. \r\nVESSEL
) similar to the prior and may represent a s
sions\r\nSOFT TISSUES: Within normal limit
SIDENT/FELLOW: Ruchika Podury, MD'
```

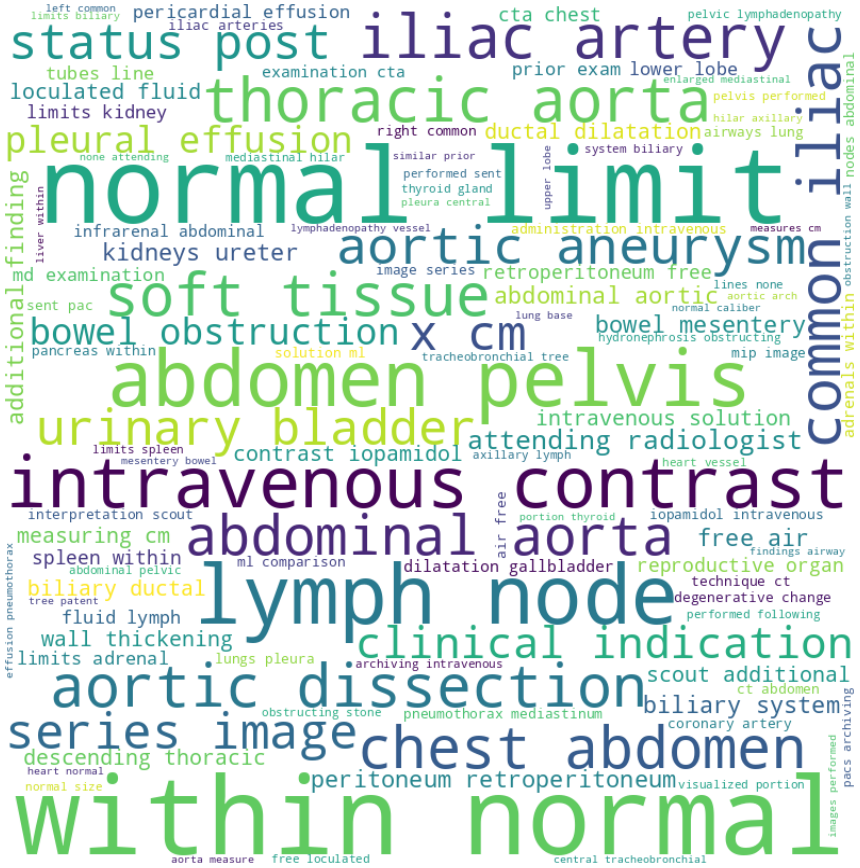
```
clean_text(df_new.Text.iloc[0])
```

```
'examination cta aortic dissection chest abdomen pelvis including iliac arteries impression examination limited lack noncontrast ct scan abdo
men pelvis status post distal aorta biiliac endovascular stenting abdominal aorta aneurysm measuring cm significantly increased size prior ex
am several areas increased attenuation within aneurysm sac detailed suggestive presence endoleak possibly type though evaluation limited due
lack noncontrast images abdomen pelvis sma ostial noncalcified plaque thrombus causing moderate stenosis thick walled gallbladder likely cont
racted small hypodense areas within proximal pancreatic body tail evaluated elective mri abdomen clinical indication abd pain aaa us aaa abdo
minal aortic aneurysm technique ct scanning chest performed without administration intravenous contrast following cta chest abdomen pelvis pe
rformed following intravenous administration contrast reformats constructed sagittal coronal planes mip images performed sent pacs archiving
intravenous contrast iopamidol intravenous solution ml comparison prior ct chest abdomen pelvis dated endovascular ultrasound dated interpret
ation limited streak artifacts scout additional findings airways lungs pleura central tracheobronchial tree patent bibasilar linear atelectas
is scarring small right pleural effusion pneumothorax mediastinum enlarged mediastinal hilar axillary lymph nodes visualized portion thyroid
gland heterogeneous small nodules calcifications heart vessels cardiomegaly coronary artery aortic calcifications thoracic aorta normal calib
er small pericardial effusion liver within normal limits biliary system biliary ductal dilatation gallbladder cholelithiasis gallbladder wall
nonspecifically thickened probably secondary contraction pancreas ill defined hypodensity proximal pancreatic body tail measuring cm series i
mage spleen within normal limits adrenals bilateral adrenal gland thickening kidneys ureters bilateral atrophic kidneys multiple cysts measur
ing cm right lower pole hydronephrosis bowel mesentery bowel obstruction wall thickening appendix normal urinary bladder underdistended limit
ing evaluation reproductive organs pelvis partially obscured streak artifact left hip prosthesis peritoneum retroperitoneum trace free pelvic
fluid lymph nodes abdominal pelvic lymphadenopathy vessels patient status post distal aorta biiliac endovascular stenting x cm inferior abdom
inal aorta aneurysm significantly increased size prior exam x cm although active extravasation noted current exam several areas increased att
enuation within aneurysm sac example series image series image series image noncalcified plaque thrombus ostium sma causing moderate stenosis
bones prior left hip replacement mild scoliosis lumbar spine diffuse osteopenia noted soft tissues within normal limits tubes lines none atte
nding radiologist bindu kaul md resident fellow andrea furlani md'
```

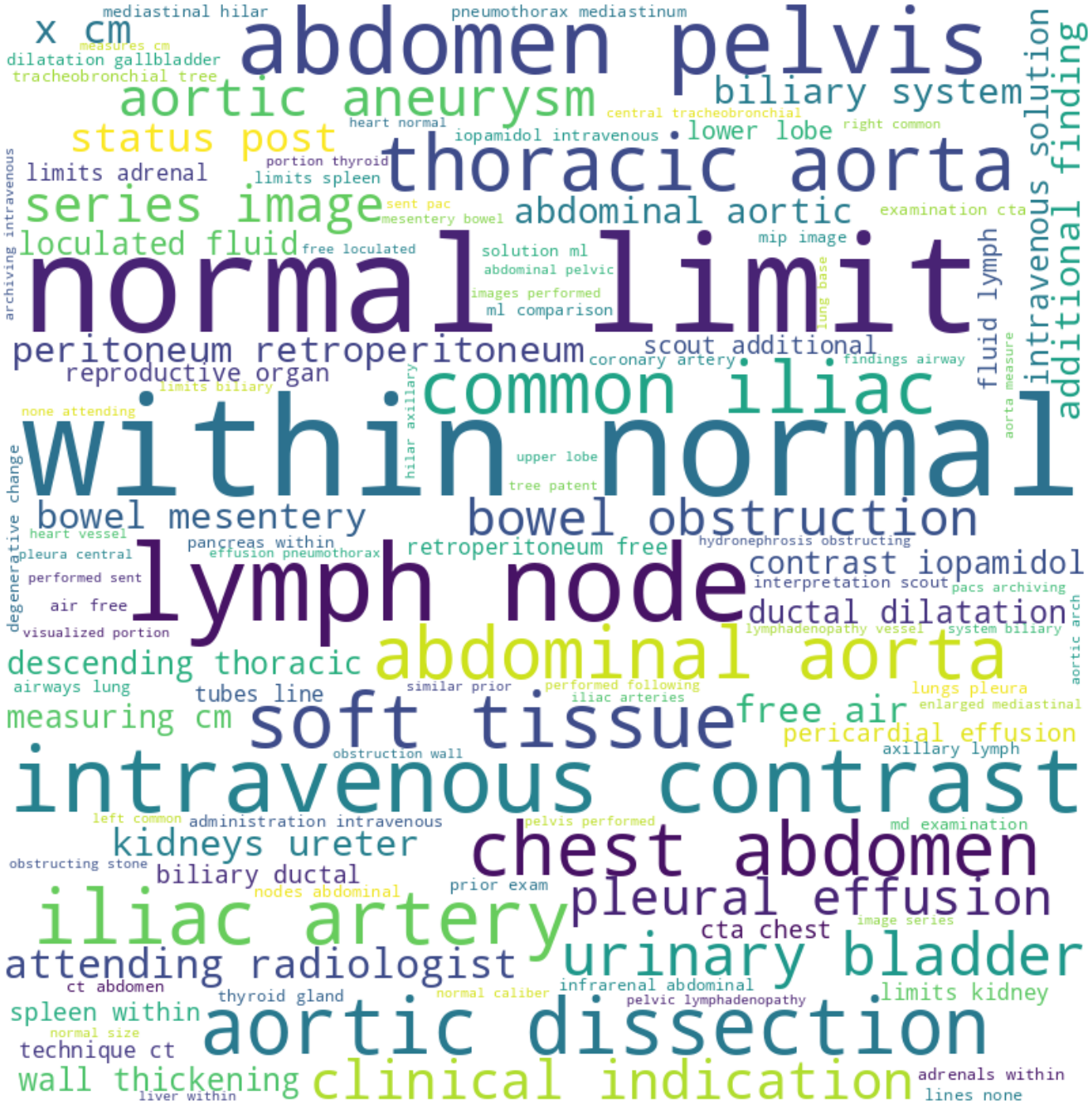
Abdelrahman Mohamed

Word Cloud

Important Keywords for True Values



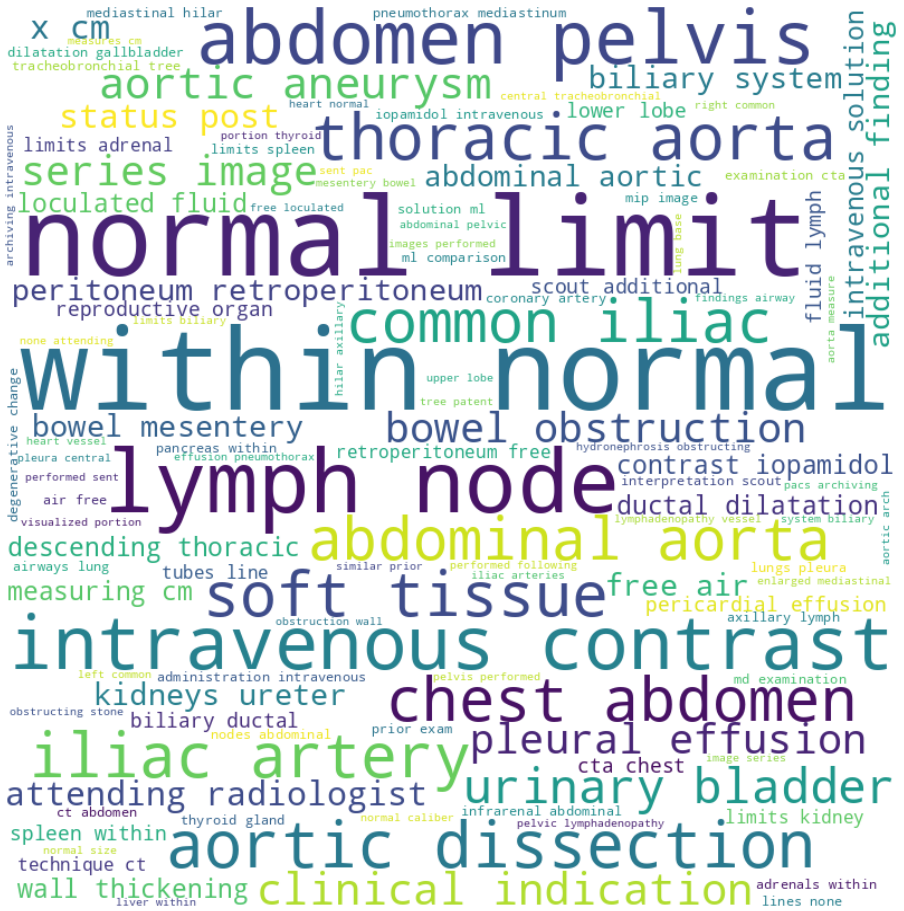
Important Keywords for False Values



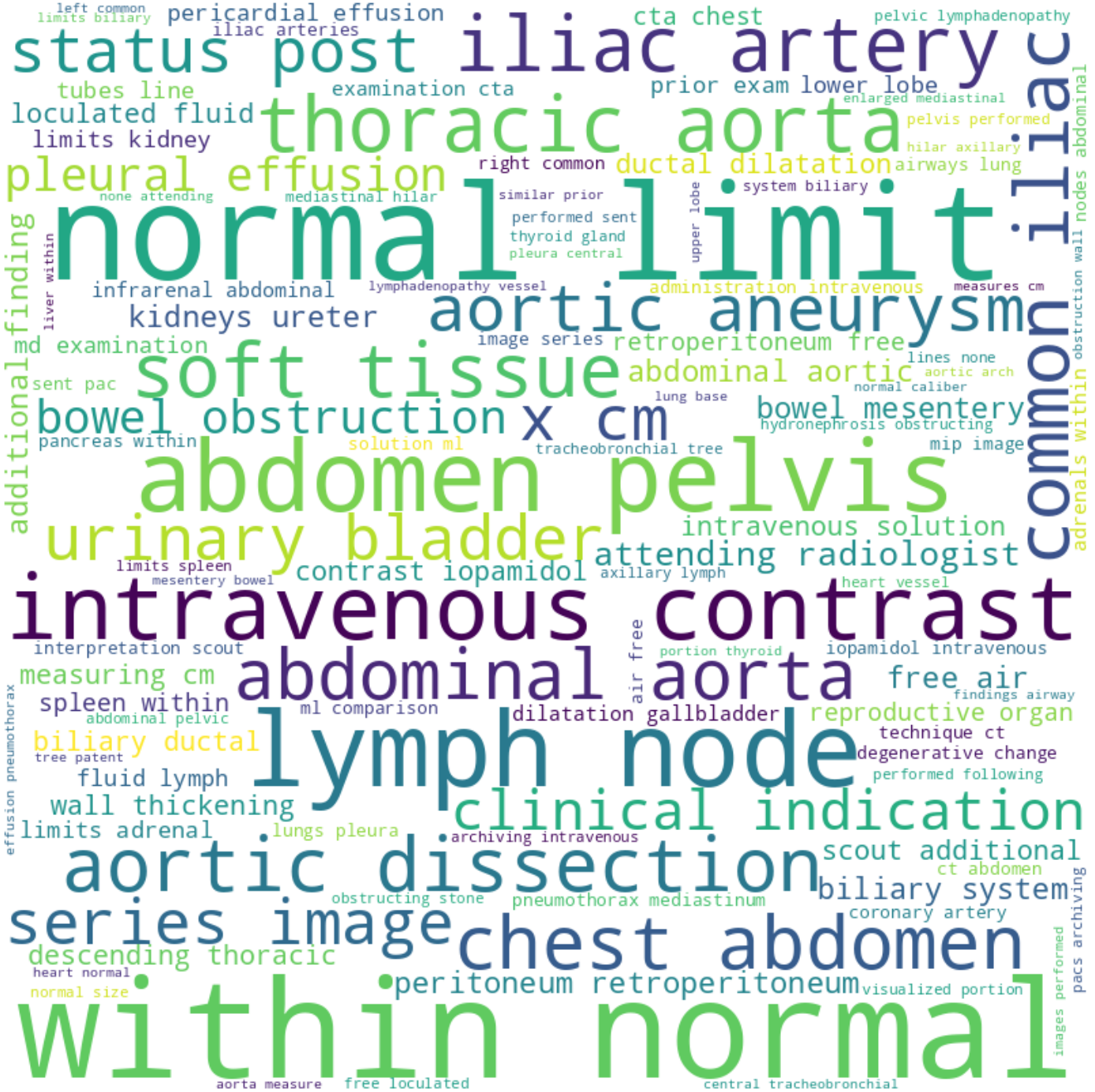
Note: Biggest words are most frequent and smaller ones are less frequent

Word Cloud

Important Keywords for False Values



Important Keywords for True Values



Note: Biggest words are most frequent and smaller ones are less frequent

Preprocessing

Tokenization and convert features into sequences of numbers

```
X
```

```
array([[ 82,  34,   7, ...,  38,   6, 365],
       [ 82,  34,   7, ...,   3,  11,  64],
       [ 82,  34,  21, ...,  30, 150,  12],
       ...,
       [ 82,  14,  16, ...,   0,   0,   0],
       [ 82,  34,   7, ...,   0,   0,   0],
       [ 82,  34,   7, ...,   0,   0,   0]], dtype=int32)
```

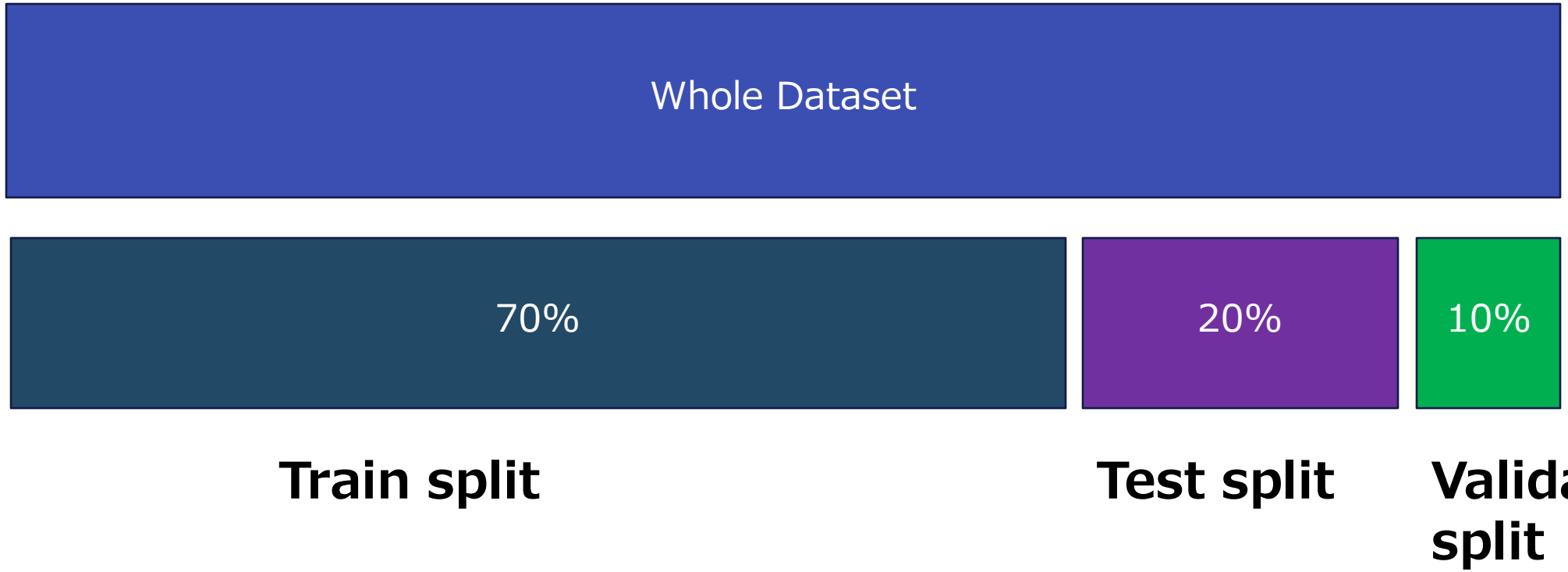
Sequence Length: 300
Padding Method: post

```
y
```

```
array([1, 1, 1, ..., 0, 0, 0])
```

Labels enconding
Give 0 to False values and 1 to True Values

Train/Test Split



Model Architecture

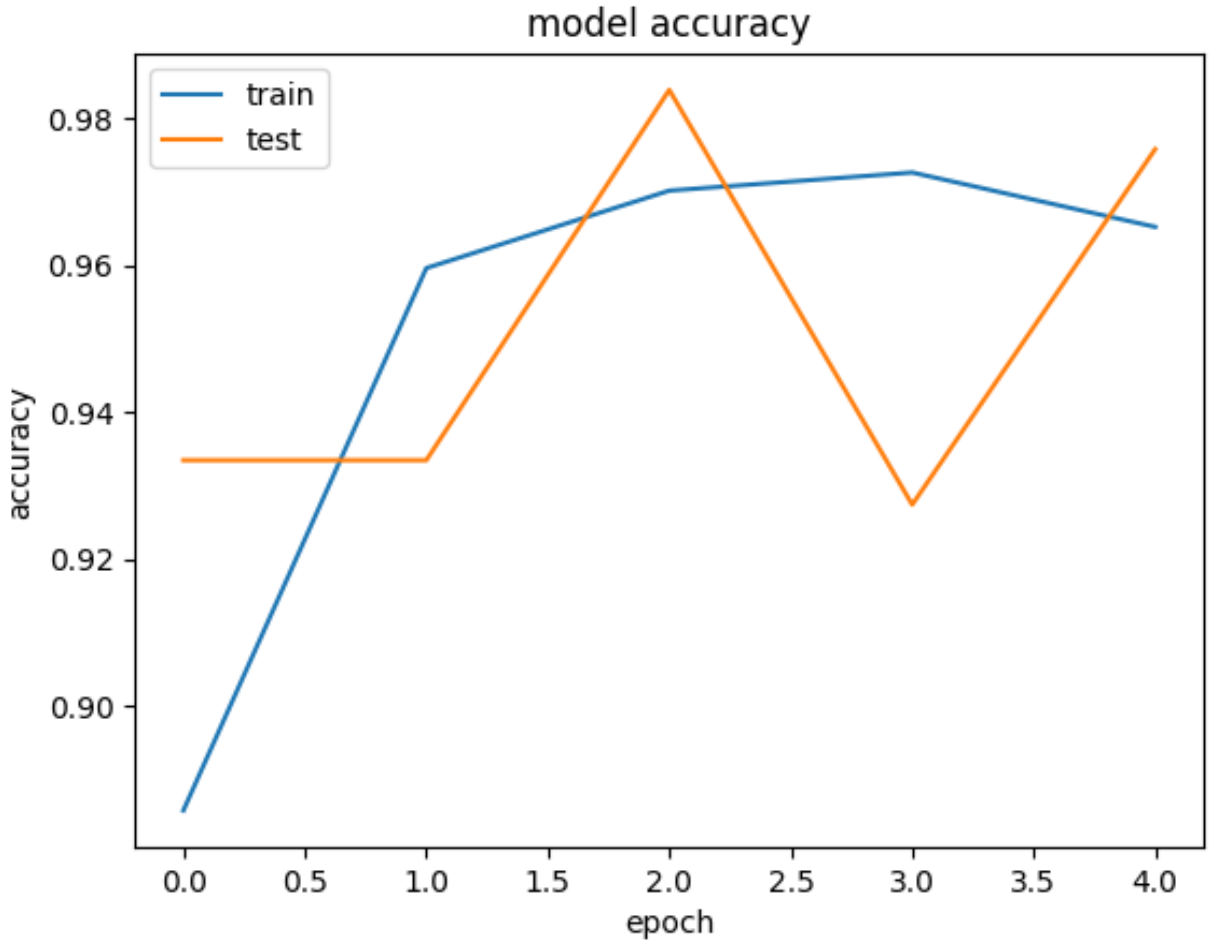
Used Architecture

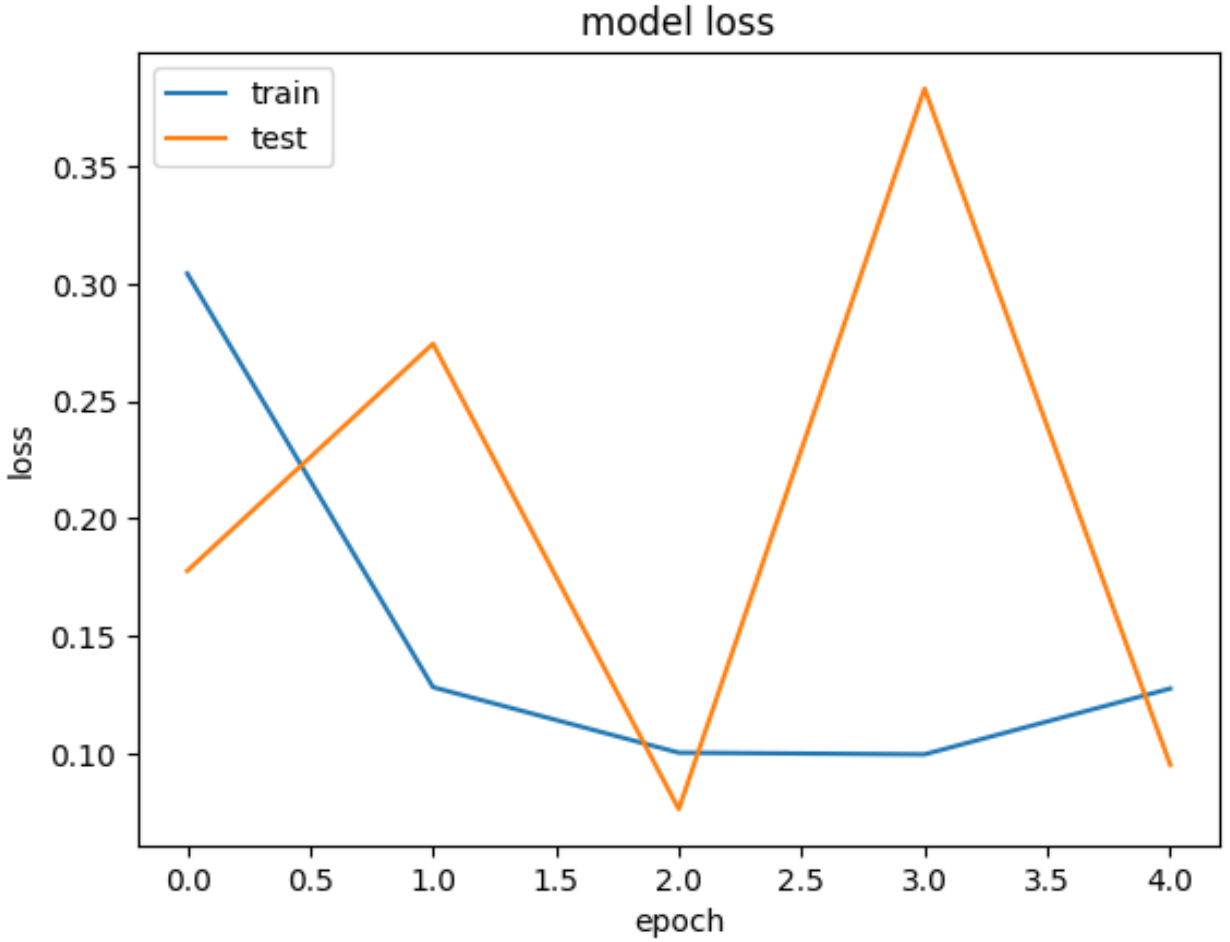
```
model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, 300, 300)          1800000

 lstm_2 (LSTM)               (None, 300, 100)          160400

 dense_3 (Dense)             (None, 300, 100)          10100

 lstm_3 (LSTM)               (None, 32)                17024

 dense_4 (Dense)             (None, 32)                1056

 dense_5 (Dense)             (None, 1)                 33

=================================================================
Total params: 1988613 (7.59 MB)
Trainable params: 188613 (736.77 KB)
Non-trainable params: 1800000 (6.87 MB)
```

Model Training

Epochs: 5





Abdelrahman Mohamed

Model Evaluation

Model Score on test data

Loss function: Binary Cross-Entropy

Loss: 0.1

Accuracy: 97%

```
ev = model.evaluate(X_test,y_test)

print(f"Loss: {round(ev[0], 2)}, Accuracy: {round(ev[1], 2)}")
```

```
18/18 [==============================] - 2s 110ms/step - loss: 0.0988 - accuracy: 0.9746
Loss: 0.1, Accuracy: 0.97
```

Thank You

Abdelrahman Mohamed