Overall, the folder contains 28 code files and 43 data files. Their names, content, objective and related figures are illustrated in below Table.

In terms of directory, the folder names and their files are as below:

Code:

- Random sampling - Inception - 5000 (random 0-4): 5 notebooks
    - the original framework (random split) with 5000 samples and change in random states to check on stability
    - the one with random state = 0 contains the comparison between 'one-to-one' and 'one-to-all'
- Stratified sampling - Inception - 5000 (random 0-4): 5 notebooks
    - stratified framework with topic modelling section with 5000 sample and change random states
- Random sampling - Inception - 1000-9000-13000 (random =0): 3 notebooks
    - the original framework (random split) with 1000, 9000 and 13000 samples
- Stratified sampling - Inception - 1000-9000-13000 (random=0): 3 notebooks
    - stratified framework with topic modelling section with 1000, 9000, 13000 samples
- Stratified sampling - Compared CNNs - 5000 (epoch 20-30-100): 8 notebooks:
    - Stratified framework with 5000 samples
    - VGG and EffNet: 3 files, epoch = 20, 30, 100
    - Inception: 2 file, epoch = 30, 100
- Topic modelling (Merged- Unmerged- Different sample): 3 notebooks:
    - define appropriate number of method and topic including merged and unmerged routes, and 1000, 5000, 9000 and 13000 samples
- Result - Discussion.ipynb: The evaluation file

Data:

- similar sub-folders with data files generated from respective code files as above
- 3 types:
    - BLEU score
    - prediction (set of image path, BLEU and prediction)
    - Topic: image_path + original text + assigned topic

Objective	Code file	Data file	Sampling	Size	Random	Encoder	Epoch	Data Content	Figure
Comparison & Analysis	Result - Discussion.ipynb	This file contains most of the analysis and figures showed in the Dissertation. There are 4 sections following 4 objectives. Each section loads the .csv file listed below for analysis							Figure: 4.1, 4.4, 4.5, 5.3 – 5.10 Table: 4.1
Evaluation metric	Without TM - Inception - 5k (rd=0).ipynb	'withoutTM-bleus oto 5000 (rd0).csv'	Random	5000	0	InceptionV3	20	BLEU one-to-one	Figure: 5.1
		'withoutTM-bleus ota 5000 (rd0).csv'	Random	5000	0	InceptionV3	20	BLEU one-to-all	
		'withoutTM-pred 5000 (rd0).csv'	Random	5000	0	InceptionV3	20	Prediction + Image path one-to-all	
Topic modelling	Unmerged cap - LDA_vs_NMF - 5k.ipynb	N/A	N/A	5000	N/A	N/A	N/A	Coherence score & Empty cap	Figure: 4.2 (right), 4.3 (right)
	Merged cap - LDA_vs_NMF - 5k.ipynb	N/A	N/A	5000	N/A	N/A	N/A	Coherence score & Empty cap	Figure: 4.2 (left), 4.3 (left)
	Merged cap - LDA_vs_NMF - Sample change.ipynb	N/A	N/A	1000, 9000, 13000	N/A	N/A	N/A	Coherence score & Empty cap	Figure: 5.2
Stratification Impact: Accuracy	Without TM - Inception - 5k (rd=1).ipynb	'withoutTM-bleus 5000 (rd1).csv'	Random	5000	1	InceptionV3	20	BLEU one-to-all	N/A
		'withoutTM-pred 5000 (rd1).csv'	Random	5000	1	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Without TM - Inception - 5k (rd2).ipynb	'withoutTM-bleus 5000 (rd2).csv'	Random	5000	2	InceptionV3	20	BLEU one-to-all	N/A

		'withoutTM-pred 5000 (rd2).csv'	Random	5000	2	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Without TM - Inception - 5k (rd=3).ipynb	'withoutTM-bleus 5000 (rd3).csv'	Random	5000	3	InceptionV3	20	BLEU one-to-all	N/A
		'withoutTM-pred 5000 (rd3).csv'	Random	5000	3	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Without TM - Inception - 5k (rd=4).ipynb	'withoutTM-bleus 5000 (rd4).csv'	Random	5000	4	InceptionV3	20	BLEU one-to-all	N/A
		'withoutTM-pred 5000 (rd4).csv'	Random	5000	4	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 5k (rd=0) .ipynb	'merged-bleus 5000 (rd0).csv'	Stratified	5000	0	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 5000 (rd0).csv'	Stratified	5000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
		'5000wtopic.csv'	Stratified	5000	0	InceptionV3	20	Images with topic	N/A
	Merged- With TM - Inception - 5k (rd=1).ipynb	'merged-bleus 5000 (rd1).csv'	Stratified	5000	1	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 5000 (rd1).csv'	Stratified	5000	1	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 5k (rd=2).ipynb	'merged-bleus 5000 (rd2).csv'	Stratified	5000	2	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 5000 (rd2).csv'	Stratified	5000	2	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 5k (rd=3).ipynb	'merged-bleus 5000 (rd3).csv'	Stratified	5000	3	InceptionV3	20	BLEU one-to-all	N/A

		'merged-pred 5000 (rd3).csv'	Stratified	5000	3	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 5k (rd=4).ipynb	'merged-bleus 5000 (rd4).csv'	Stratified	5000	4	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 5000 (rd4).csv'	Stratified	5000	4	InceptionV3	20	Prediction + Image path one-to-all	N/A
Stratification impact: Scalability	Without TM - Inception - 1k.ipynb	'withoutTM-bleus 1000.csv'	Random	1000	0	InceptionV3	20	BLEU one-to-all	N/A
		'withoutTM-pred 1000.csv'	Random	1000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 1k.ipynb	'merged-bleus 1000.csv'	Stratified	1000	0	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 1000.csv'	Stratified	1000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
		'1000wtopic.csv'	Stratified	1000	0	InceptionV3	20	Images with topic	N/A
	Without TM - Inception - 9k.ipynb	'withoutTM-bleus 9000.csv'	Random	9000	0	InceptionV3	20	BLEU one-to-all	N/A
		'withoutTM-pred 9000.csv'	Random	9000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 9k.ipynb	'merged-bleus 9000.csv'	Stratified	9000	0	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 9000.csv'	Stratified	9000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
		'9000wtopic.csv'	Stratified	9000	0	InceptionV3	20	Images with topic	N/A

	Without TM - Inception - 13k.ipynb	'withoutTM-bleus 13000.csv'	Random	13000	0	InceptionV3	20	BLEU one-to-all	N/A
		'withoutTM-pred 13000.csv'	Random	13000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
	Merged- With TM - Inception - 13k.ipynb	'merged-bleus 13000.csv'	Stratified	13000	0	InceptionV3	20	BLEU one-to-all	N/A
		'merged-pred 13000.csv'	Stratified	13000	0	InceptionV3	20	Prediction + Image path one-to-all	N/A
		'13000wtopic.csv'	Stratified	13000	0	InceptionV3	20	Images with topic	N/A
Image Encoders	Merged- With TM - Inception - 5k (rd=0) - epoch = 30.ipynb	'merged-bleus 5000 30epoch.csv'	Stratified	5000	0	InceptionV3	30	BLEU one-to-all	Figure: 4.7 (avg. BLEU)
		'merged-pred 5000 30epoch.csv'	Stratified	5000	0	InceptionV3	30	Prediction + Image path one-to-all	
	Merged- With TM - Inception - 5k (rd=0) - epoch = 100.ipynb	N/A	Stratified	5000	0	InceptionV3	100	N/A	Figure: 4.6 (middle), 4.7 (avg. BLEU)
	Merged - With TM - VGG - 5k (rd=0) epoch = 20.ipynb	N/A	Stratified	5000	0	VGG16	20	N/A	Figure: 4.7 (avg. BLEU)
	Merged- With TM - Inception - 5k (rd=0) - epoch = 30.ipynb	'merged - VGG bleus 30epoch.csv'	Stratified	5000	0	VGG16	30	BLEU one-to-all	Figure: 4.7 (avg. BLEU)
		'merged - VGG pred 30epoch.csv'	Stratified	5000	0	VGG16	30	Prediction + Image path one-to-all	
	Merged - With TM - VGG - 5k (rd=0) - epoch =100.ipynb	N/A	Stratified	5000	0	VGG16	100	N/A	Figure: 4.6 (left), 4.7 (avg. BLEU)

	Merged - With TM - EfficientNetB7 - 5k (rd=0) - epoch=20.ipynb	'merged - EffNet bleus 20epoch.csv'	Stratified	5000	0	EfficientNetB7	20	BLEU one-to-all	Figure: 4.7 (avg. BLEU)
		'merged - EffNet pred 20epoch.csv'	Stratified	5000	0	EfficientNetB7	20	Prediction + Image path one-to-all	
	Merged - With TM - EfficientNetB7 - 5k (rd=0) - epoch=30 .ipynb	N/A	Stratified	5000	0	EfficientNetB7	30	N/A	Figure: 4.7 (avg. BLEU)
	Merged - With TM - EfficientNetB7 - 5k (rd=0) - epoch 100 .ipynb	N/A	Stratified	5000	0	EfficientNetB7	100	N/A	Figure: 4.6 (right), 4.7 (avg. BLEU)