

Modelling

August 25, 2023

```
Loaded dictionary: dict_keys(['class_metadata', 'features', 'labels'])
```

1 Feature Extraction

Features were extracted by generating LBP patterns from curvelet transforms of images as mentioned in <https://doi.org/10.1016/j.eswa.2016.02.019>

```
Explained Variance Ratio: [0.32174306 0.05224153 0.04271408]
```

2 Dimensionality reduction

Dimensions are reduced using PCA algorithm, into 3 principal components and are plotted equivalently.

We can see that there is a class imbalance from the figure.

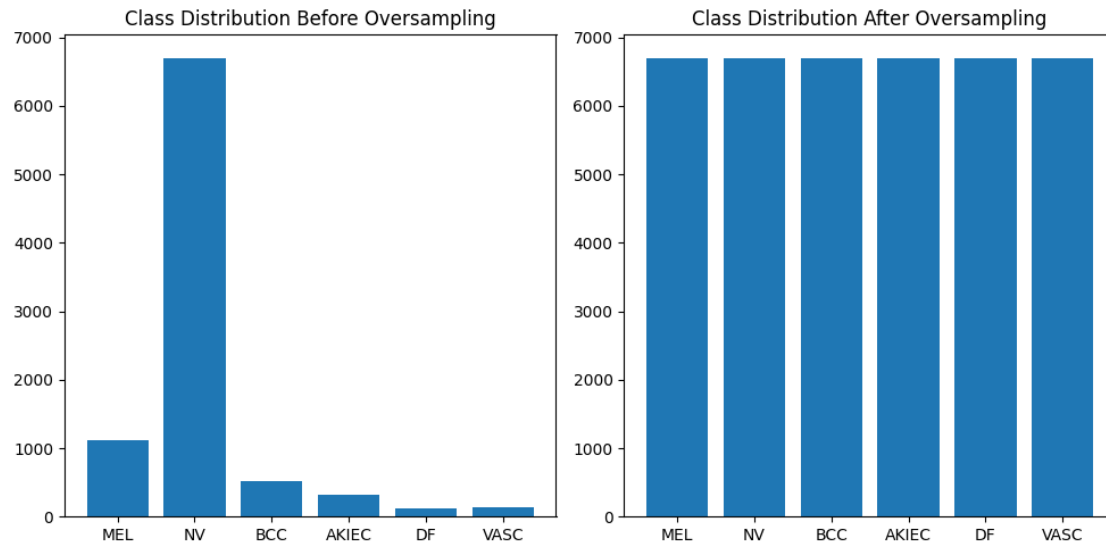


3 Class Imbalance

By leveraging SMOTE for oversampling we are able to tackle the class imbalance issue as shown below.

```
Class distribution before oversampling: Counter({'NV': 6705, 'MEL': 1113, 'BCC': 514, 'AKIEC': 327, 'VASC': 142, 'DF': 115})
```

```
Class distribution after oversampling: Counter({'MEL': 6705, 'NV': 6705, 'BCC': 6705, 'AKIEC': 6705, 'DF': 6705, 'VASC': 6705})
```



4 Modelling SVM, Random Forest, Adaboost

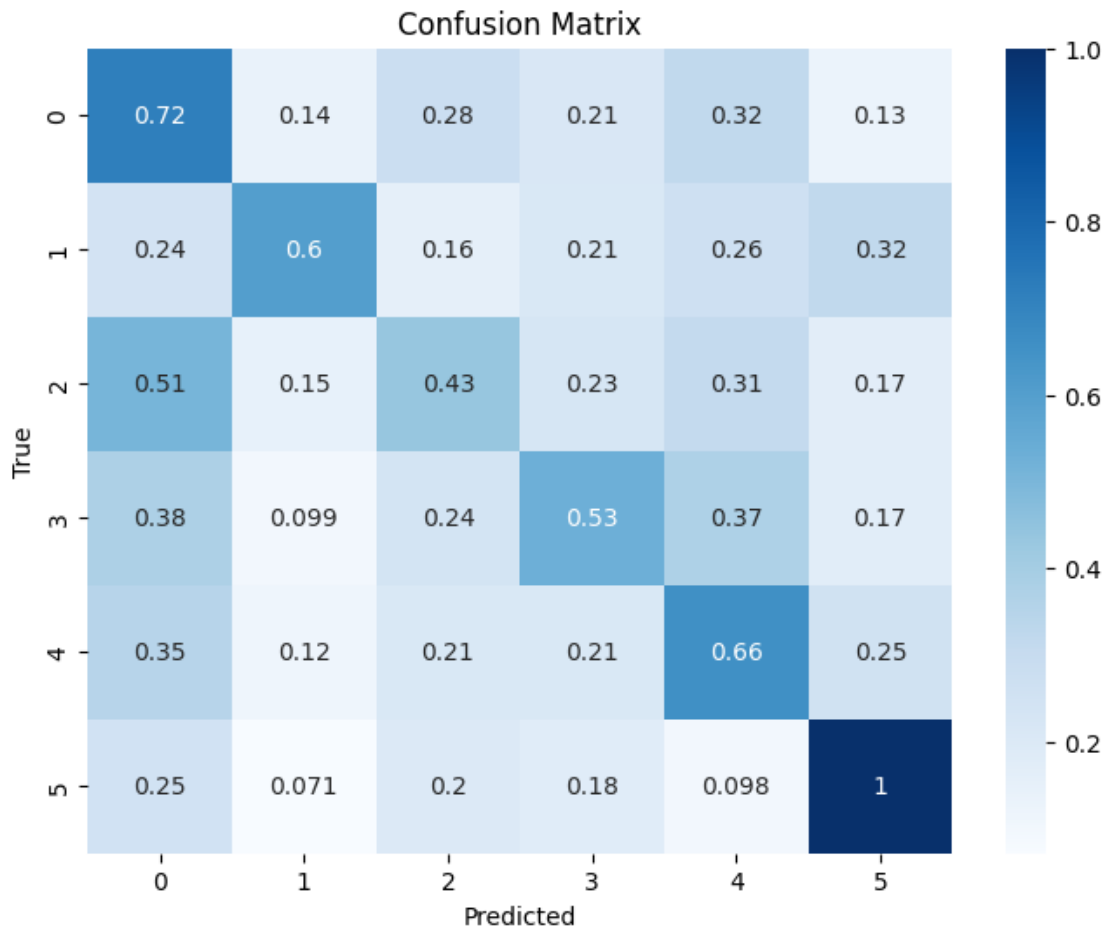
4.1 SVM

4.1.1 Overall stats of 10-fold stratified cross validation

```
[9]:
```

Fold	accuracy
0	0.376833
1	0.363908
2	0.365648
3	0.357693
4	0.365896
5	0.366890
6	0.353468
7	0.365648
8	0.372110
9	0.368133

4.1.2 Confusion matrix



Precision: 0.8128612716763006

Recall: 0.7129277566539924

F1-Score: 0.7596218771100608

Support: 3156.0

4.2 Random Forrest

4.2.1 Overall stats of 10-fold startified cross validation

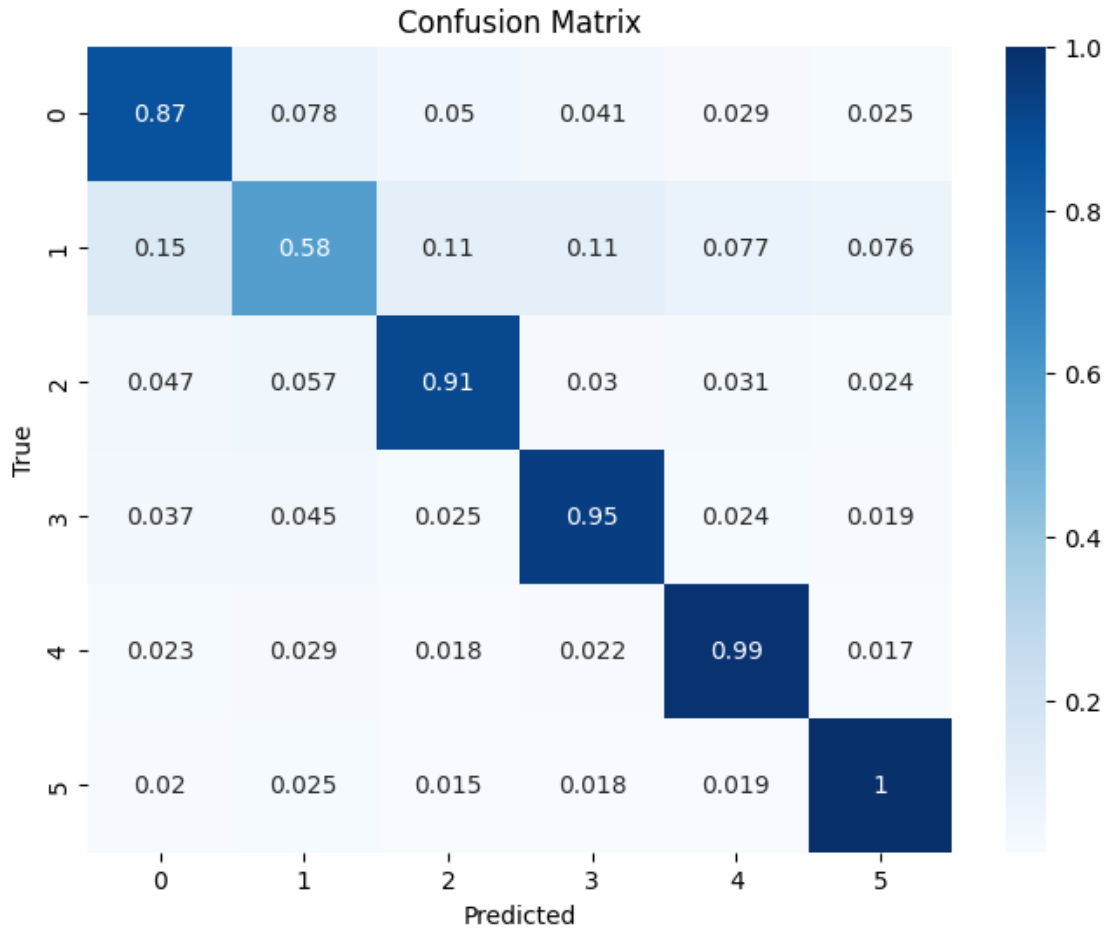
```
[12]: Fold accuracy
0     0  0.810589
1     1  0.805618
2     2  0.802635
3     3  0.801641
4     4  0.804623
5     5  0.803629
6     6  0.799901
```

```

7      7  0.805866
8      8  0.807109
9      9  0.804126

```

4.2.2 Confusion matrix



```

Precision: 0.8813137596417019
Recall: 0.7952402334979793
F1-Score: 0.8360675085565915
Support: 4454.0

```

4.3 Adaboost

4.3.1 Overall stats of 10-fold startified cross validation

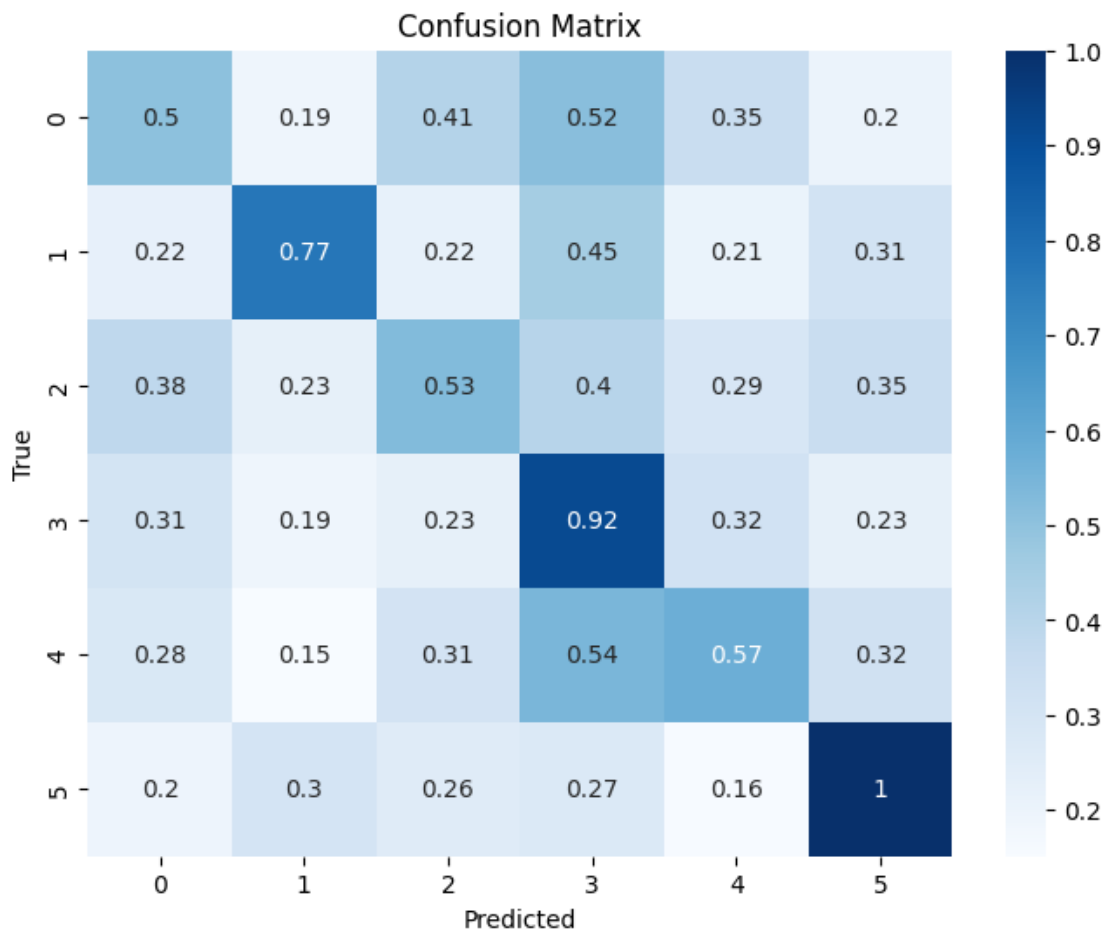
```

[15]:  Fold  accuracy
      0    0  0.338802
      1    1  0.329356

```

2	2	0.330848
3	3	0.311459
4	4	0.327119
5	5	0.322148
6	6	0.339548
7	7	0.333085
8	8	0.331096
9	9	0.318916

4.3.2 Confusion Matrix



Precision: 0.7986509274873524
 Recall: 0.7758846657929227
 F1-Score: 0.7871032075785276
 Support: 3052.0