

SVM Binary Market Classification Model for t-SNE Generated Clusters of NREL Dataset:  
*Rooftop Solar Photovoltaic Technical Potential in the United States*

TABLES & FIGURES

ChE 765 – Artificial Intelligence and Machine Learning Fundamentals

McMaster University, Hamilton, ON. Canada  
L8S 4L7

August 16, 2020  
By: Mohammad Akhtar, M.A.Sc. 2019 – 2021

## Table of Contents

List of Tables	1
List of Figures	1
1 Dimensionality Reduction and Data Visualization Summary	7
2 Principal Component Analysis, <i>PCA</i>	11
3 t-Distributed Stochastic Neighbour Embedding, <i>t-SNE</i>	13
4 Uniform Manifold Approximation and Projection, <i>UMAP</i>	33
5 t-Distributed Stochastic Neighbour Embedding, <i>t-SNE (Re-classified)</i>	51
6 Uniform Manifold Approximation and Projection, <i>UMAP (Re-classified)</i>	56

## List of Tables

Table 1. t-SNE visualisation summary: components = 3; gradient calculation method = <i>Barnes-Hut</i> ; pairwise distance metric = <i>Euclidean</i> ; learning rate = 500.	7
Table 2. UMAP visualisation summary: components = 3; spread = 1; pairwise distance metric = <i>Euclidean</i> ; learning rate = 10.	8
Table 3. t-SNE visualisation summary for <i>re-classified</i> dataset: components = 2; gradient calculation method = <i>Barnes-Hut</i> ; pairwise distance metric = <i>Euclidean</i> ; learning rate = 10.	8
Table 4. UMAP visualisation summary for <i>re-classified</i> dataset: components = 2; spread = 1; pairwise distance metric = <i>Euclidean</i> ; learning rate = 10.	9
Table 5. PCA, t-SNE, and UMAP compute times for <i>re-classified</i> dataset.	9
Table 6. t-SNE cluster SVM polynomial kernel (degree = 10) classification report.	9
Table 7. t-SNE cluster SVM polynomial kernel (degree = 10) confusion matrix.	9
Table 8. t-SNE cluster SVM radial basis function (gamma = 10) classification report.	10
Table 9. t-SNE cluster SVM radial basis function (gamma = 10) confusion matrix.	10
Table 10. t-SNE cluster SVM sigmoid classification report.	10
Table 11. t-SNE cluster SVM sigmoid confusion matrix.	10

## List of Figures

Figure 2.1. PCA visualization for original dataset. Classification labels <i>small</i> , <i>medium</i> , <i>large</i> represented by <i>red</i> , <i>green</i> , <i>blue</i> , respectively.	11
Figure 2.2. PCA visualization for <i>re-classified</i> dataset. Classification labels <i>small</i> and <i>large</i> represented by <i>blue</i> and <i>orange</i> , respectively.	12
Figure 3.1. t-SNE visualisation: components = 3; gradient calculation method = <i>Barnes-Hut</i> ; pairwise distance metric = <i>Euclidean</i> ; learning rate = 500, perplexity = 5; early exaggeration = 12; iterations = 1000. Classification labels <i>small</i> , <i>medium</i> , <i>large</i> represented by <i>red</i> , <i>green</i> , <i>blue</i> , respectively.	13
Figure 3.2. t-SNE visualisation: components = 3; gradient calculation method = <i>Barnes-Hut</i> ; pairwise distance metric = <i>Euclidean</i> ; learning rate = 500, perplexity = 30; early exaggeration = 12; iterations = 1000. Classification labels <i>small</i> , <i>medium</i> , <i>large</i> represented by <i>red</i> , <i>green</i> , <i>blue</i> , respectively.	14

Figure 3.3. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 50; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 15

Figure 3.4. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 75; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 16

Figure 3.5. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 110; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 17

Figure 3.6. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 200; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 18

Figure 3.7. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 400; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 19

Figure 3.8. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 1; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 20

Figure 3.9. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 48; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 21

Figure 3.10. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 60; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 22

Figure 3.11. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 72; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively. 23

Figure 3.12. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration =

**96**; iterations = **1000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 24

Figure 3.13. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **30**; early exaggeration = **12**; iterations = **500**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 25

Figure 3.14. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **30**; early exaggeration = **12**; iterations = **2000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 26

Figure 3.15. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **48**; iterations = **2000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 27

Figure 3.16. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **48**; iterations = **1000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 28

Figure 3.17. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **60**; iterations = **2000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 29

Figure 3.18. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **1000**, perplexity = **400**; early exaggeration = **60**; iterations = **2000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 30

Figure 3.19. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **10**, perplexity = **400**; early exaggeration = **60**; iterations = **5000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 31

Figure 3.20. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **10**, perplexity = **110**; early exaggeration = **60**; iterations = **1000**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 32

Figure 4.1. UMAP visualisation: components = 3; spread = **1**; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 2; minimum distance = **0.1**; transform queue size = **4**. Classification labels *small, medium, large* represented by *red, green, blue*, respectively. 33

Figure 4.2. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 5; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 34

Figure 4.3. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 35

Figure 4.4. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 50; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 36

Figure 4.5. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 100; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 37

Figure 4.6. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.0; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 38

Figure 4.7. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.25; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 39

Figure 4.8. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.50; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 40

Figure 4.9. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.80; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 41

Figure 4.10. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.99; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 42

Figure 4.11. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 50. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 43

Figure 4.12. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 75. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 44

Figure 4.13. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 100. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 45

Figure 4.14. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 200. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 46

Figure 4.15. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 400. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 47

Figure 4.16. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 2; minimum distance = 0.0; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 48

Figure 4.17. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 100; minimum distance = 0.99; transform queue size = 400. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 49

Figure 4.18. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 50; minimum distance = 0.50; transform queue size = 200. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively. 50

Figure 5.1. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 10, perplexity = 400; early exaggeration = 60; iterations = 5000. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 51

Figure 5.2. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 10, perplexity = 110; early exaggeration = 60; iterations = 5000. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 52

Figure 5.3. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 100, perplexity = 400; early exaggeration = 60; iterations = 5000. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 53

Figure 5.4. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 50, perplexity = 400; early exaggeration = 60; iterations = 5000. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 54

Figure 5.5. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 50, perplexity = 400; early exaggeration = 60; iterations = 5000; Barnes-Hut  $\theta$  = 0.0. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 55

Figure 6.1. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 100; minimum distance = 0.50; transform queue size = 200. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 56

Figure 6.2. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **100**; minimum distance = **0.10**; transform queue size = **400**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 57

Figure 6.3. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **50**; minimum distance = **0.10**; transform queue size = **400**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 58

Figure 6.4. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **50**; minimum distance = **0.05**; transform queue size = **800**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 59

Figure 6.5. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **50**; minimum distance = **0.50**; transform queue size = **800**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively. 60

## 1 Dimensionality Reduction and Data Visualization Summary

Table 1. t-SNE visualisation summary: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500.

Figure	Inputs			Outputs		
	Perplexity	Early Exaggeration (250 Iterations)	Maximum Iterations	KL Divergence (Early Exaggeration)	KL Divergence (Maximum Iterations)	Compute Time (s)
3.1	5	12	1000	84.13	0.9448	268.3
<b>3.2</b>	<b>30</b>	<b>12</b>	<b>1000</b>	<b>71.30</b>	<b>0.7977</b>	<b>206.0</b>
3.3	50	12	1000	68.53	0.7296	217.6
3.4	75	12	1000	66.20	0.6642	223.1
3.5	110	12	1000	64.15	0.6003	230.0
3.6	200	12	1000	60.26	0.4844	274.7
3.7	400	12	1000	55.78	0.3486	330.8
3.8	30	1	1000	1.404	1.0482	220.11
3.9	30	48	1000	410.4	0.8050	268.9
3.10	30	60	1000	540.8	0.8014	303.4
3.11	30	72	1000	670.3	0.8218	391.0
3.12	30	96	1000	1063	1.504	360.7
3.13	30	12	500	71.08	0.9226	106.3
3.14	30	12	2000	71.05	0.7754	419.2
3.15	400	48	2000	330.6	0.3520	462.3
3.16	400	48	1000	330.4	0.3496	447.1
3.17	400	60	2000	435.7	0.3519	489.5
3.18	400	60 (LR <sup>1</sup> 1000)	2000	578.9	1.614	1230
3.19	400	60 (LR10)	5000	443.5	0.3450	1642
3.20	110	60 (LR10)	1000	499.14	0.8484	258.8

---

<sup>1</sup> Learning rate (LR) varied.

Table 2. UMAP visualisation summary: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10.

Figure	Size of Local Neighbourhood	Effective Minimum Distance Between Embedded Points	Transform Queue Size	Compute Time (s)
4.1	2	0.1	4	19.52
4.2	5	0.1	4	13.91
<b>4.3</b>	<b>15</b>	<b>0.1</b>	<b>4</b>	<b>13.93</b>
4.4	50	0.1	4	16.46
4.5	100	0.1	4	23.88
4.6	15	0.0	4	13.81
4.7	15	0.25	4	13.71
4.8	15	0.50	4	13.81
4.9	15	0.8	4	14.73
4.10	15	0.99	4	13.97
4.11	15	0.1	50	16.88
4.12	15	0.1	75	14.48
4.13	15	0.1	100	16.13
4.14	15	0.1	200	13.91
4.15	15	0.1	400	13.78
4.16	2	0.0	4	20.43
4.17	100	0.99	400	21.77
4.18	50	0.50	200	16.64

Table 3. t-SNE visualisation summary for *re-classified* dataset: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 10.

Figure	Inputs			Outputs		
	Perplexity	Early Exaggeration (250 Iterations)	Maximum Iterations	KL Divergence (Early Exaggeration)	KL Divergence (Maximum Iterations)	Compute Time (s)
5.1	400	60	5000	419.8	0.3557	498.5
5.2	110	60	5000	468.3	0.6261	221.4
5.3	400	60 (LR100)	5000	418.1	0.3547	458.0
<b>5.4</b>	<b>400</b>	<b>60 (LR50)</b>	<b>5000</b>	<b>417.52</b>	<b>0.3542</b>	<b>484.0</b>
5.5	400	60 (LR50; $\theta^2=0$ )	5000	411.51	0.3616	2027

<sup>2</sup> Barnes-Hut angular size of distant node as measured from a point.

Table 4. UMAP visualisation summary for *re-classified* dataset: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10.

Figure	Size of Local Neighbourhood	Effective Minimum Distance Between Embedded Points	Transform Queue Size	Compute Time (s)
6.1	100	0.50	200	21.32
6.2	100	0.10	400	26.40
6.3	50	0.10	400	18.05
6.4	50	0.05	800	18.41
<b>6.5</b>	<b>50</b>	<b>0.50</b>	<b>800</b>	<b>18.40</b>

Table 5. PCA, t-SNE, and UMAP compute times for *re-classified* dataset.

Dimension Reduction and Visualization Algorithm	Compute Time (s)
PCA	0.07000
t-SNE	484.0
UMAP	18.40

Table 6. t-SNE cluster SVM polynomial kernel (degree = 10) classification report.

Label	Precision	Recall	F1-Score	Support
Large	0.90	0.25	0.39	768
Small	0.58	0.97	0.73	832
Accuracy	-	-	0.62	1600
Macro Avg.	0.74	0.61	0.56	1600
Weighted Avg.	0.74	0.62	0.57	1600

Table 7. t-SNE cluster SVM polynomial kernel (degree = 10) confusion matrix.

n = 1600	Predicted: Large	Predicted: Small
Actual: Large	189	579
Actual: Small	21	811

Table 8. t-SNE cluster SVM radial basis function classification report.

Label	Precision	Recall	F1-Score	Support
Large	0.99	0.99	0.99	768
Small	0.99	0.99	0.99	832
Accuracy	-	-	0.99	1600
Macro Avg.	0.99	0.99	0.99	1600
Weighted Avg.	0.99	0.99	0.99	1600

Table 9. t-SNE cluster SVM radial basis function confusion matrix.

$n = 1600$	Predicted: Large	Predicted: Small
Actual: Large	760	8
Actual: Small	8	824

Table 10. t-SNE cluster SVM sigmoid classification report.

Label	Precision	Recall	F1-Score	Support
Large	0.98	0.97	0.97	768
Small	0.97	0.98	0.98	832
Accuracy	-	-	0.97	1600
Macro Avg.	0.98	0.97	0.97	1600
Weighted Avg.	0.98	0.97	0.97	1600

Table 11. t-SNE cluster SVM sigmoid confusion matrix.

$n = 1600$	Predicted: Large	Predicted: Small
Actual: Large	742	26
Actual: Small	14	818

## 2 Principal Component Analysis, *PCA*

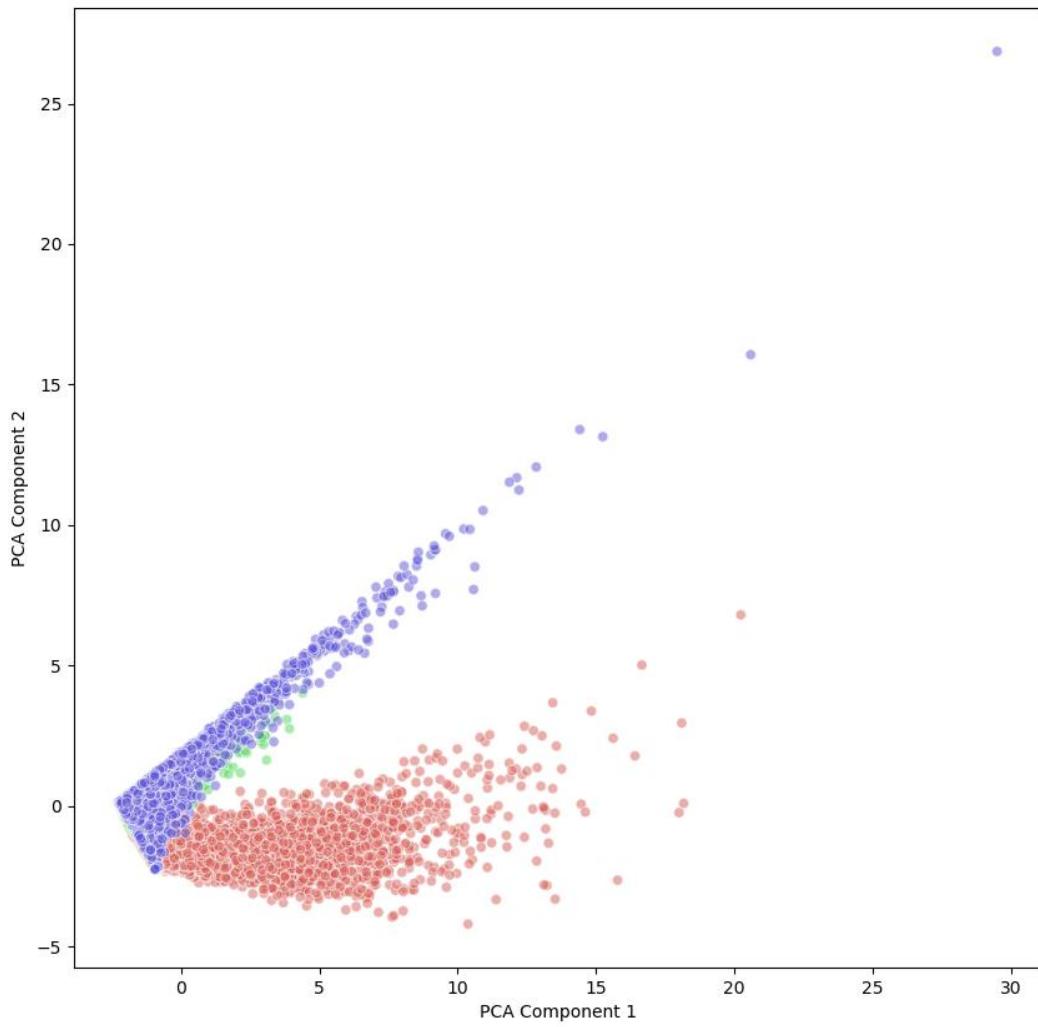


Figure 2.1. PCA visualization for original dataset. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

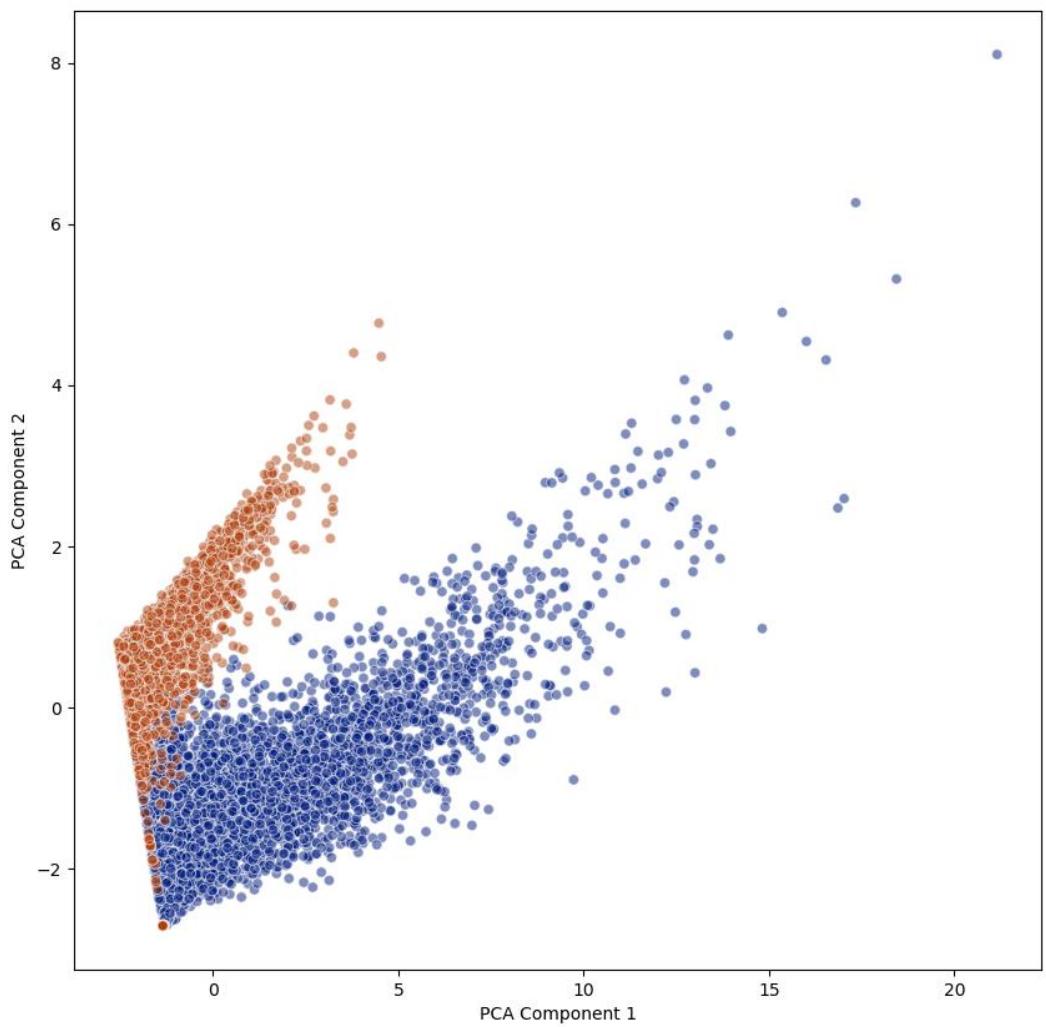


Figure 2.2. PCA visualization for *re-classified* dataset. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

### 3 t-Distributed Stochastic Neighbour Embedding, *t-SNE*

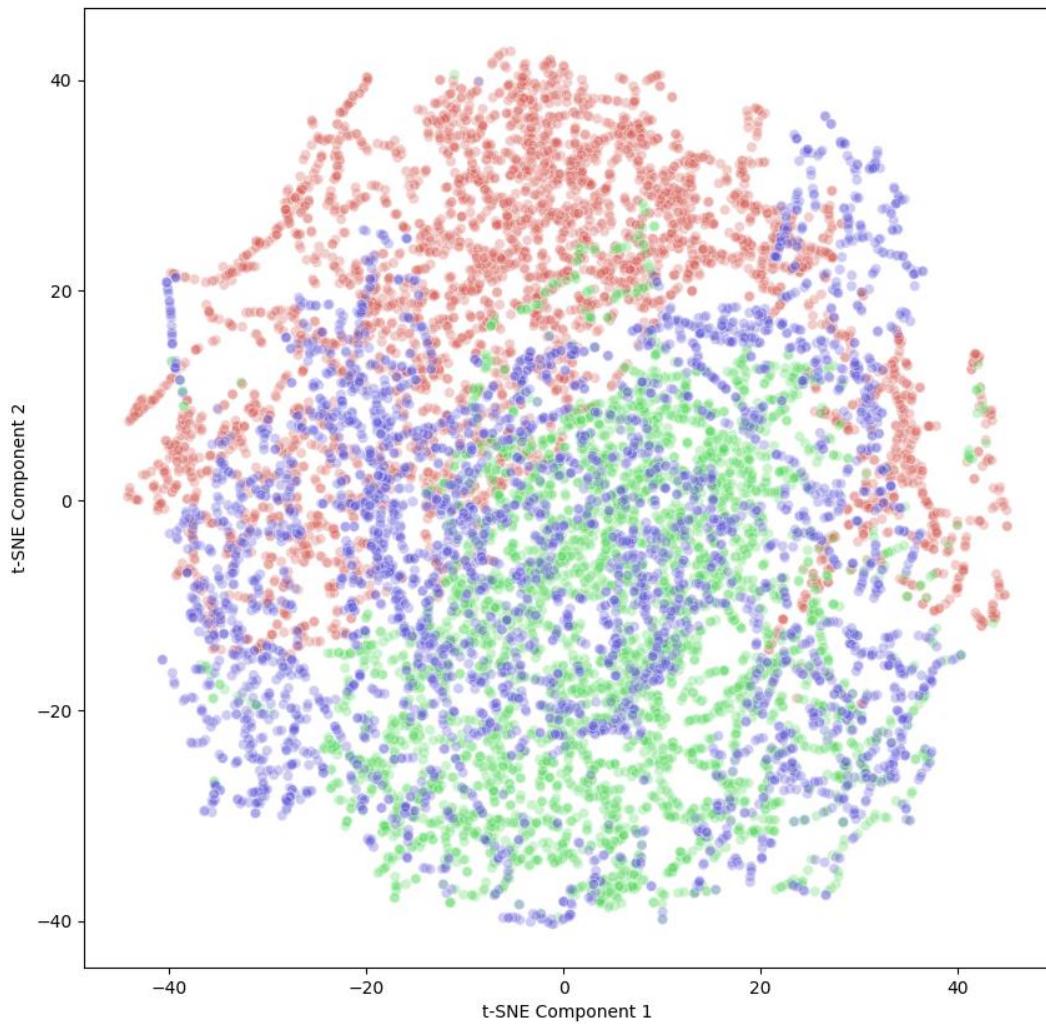


Figure 3.1. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 5; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by red, green, blue, respectively.

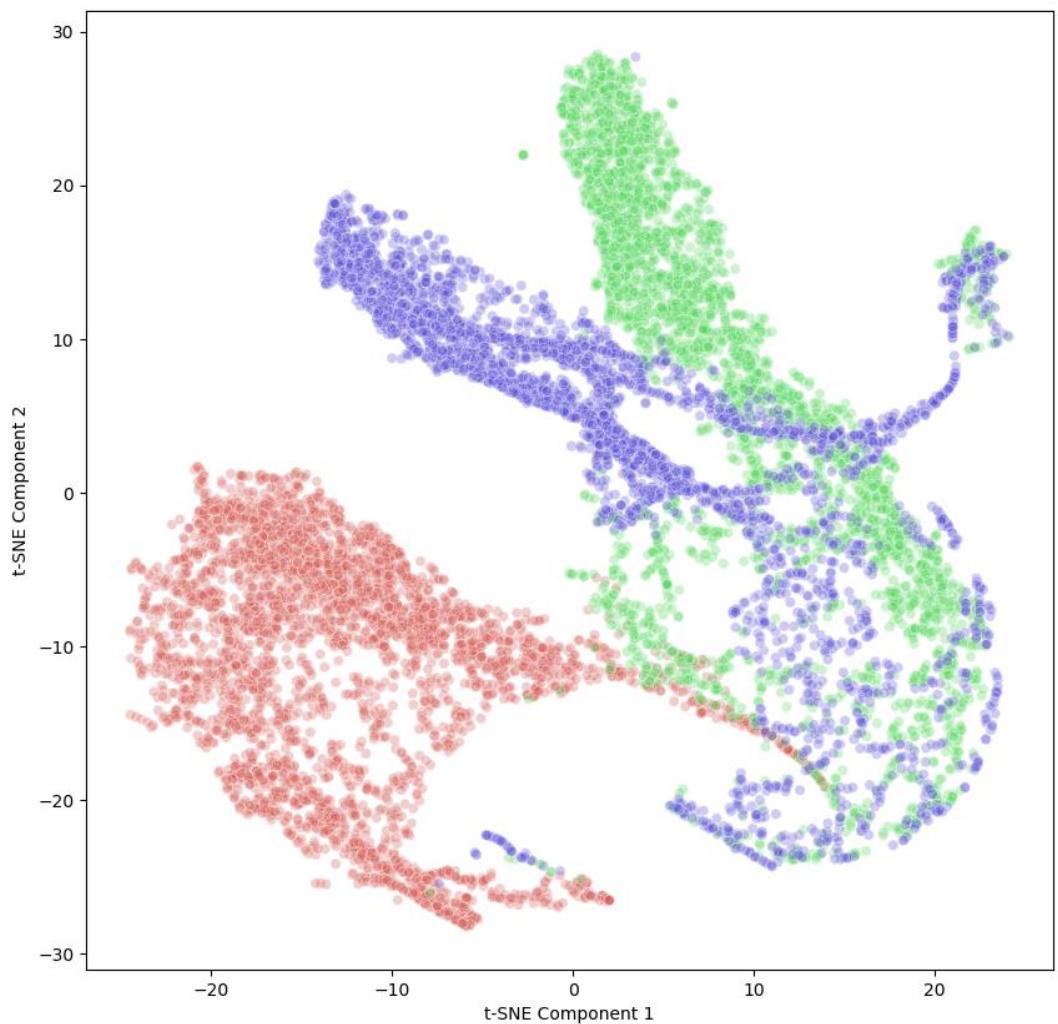


Figure 3.2. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

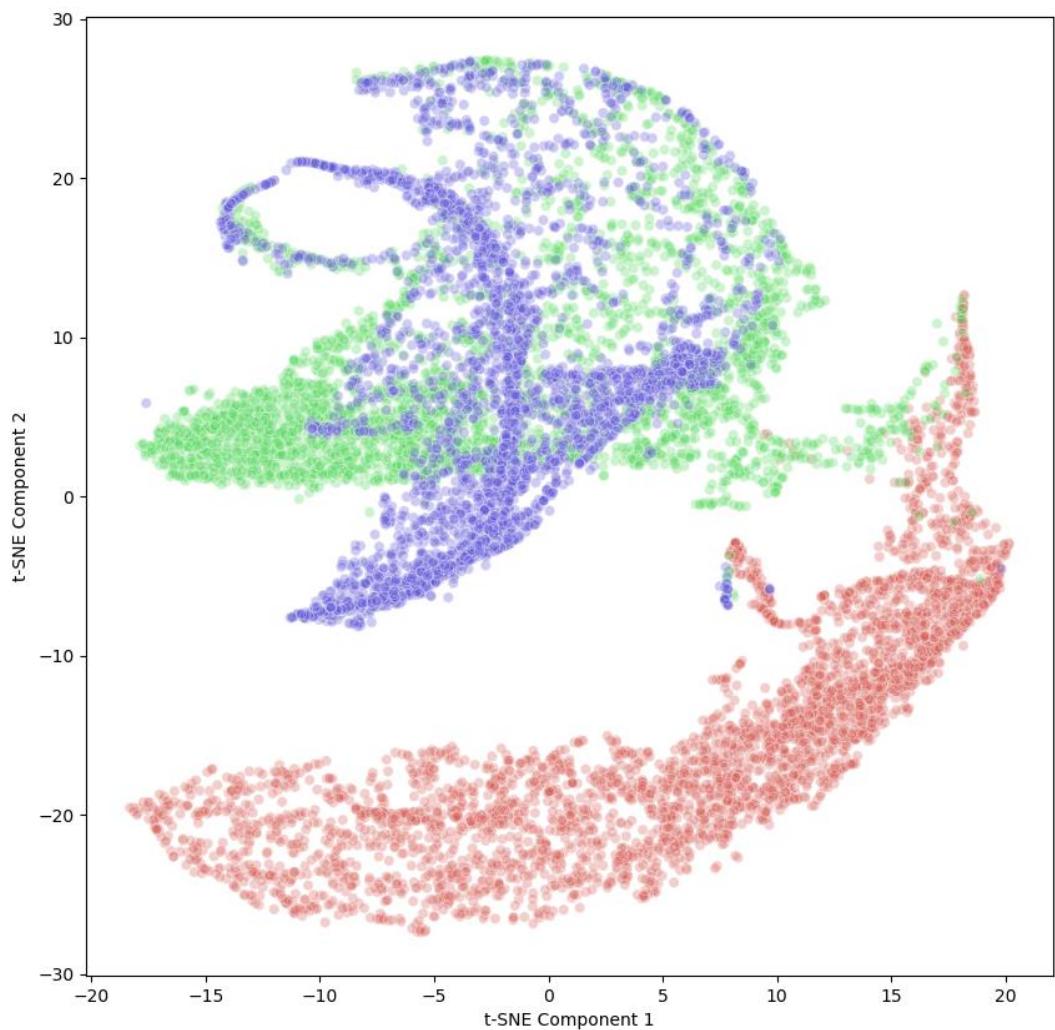


Figure 3.3. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 50; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

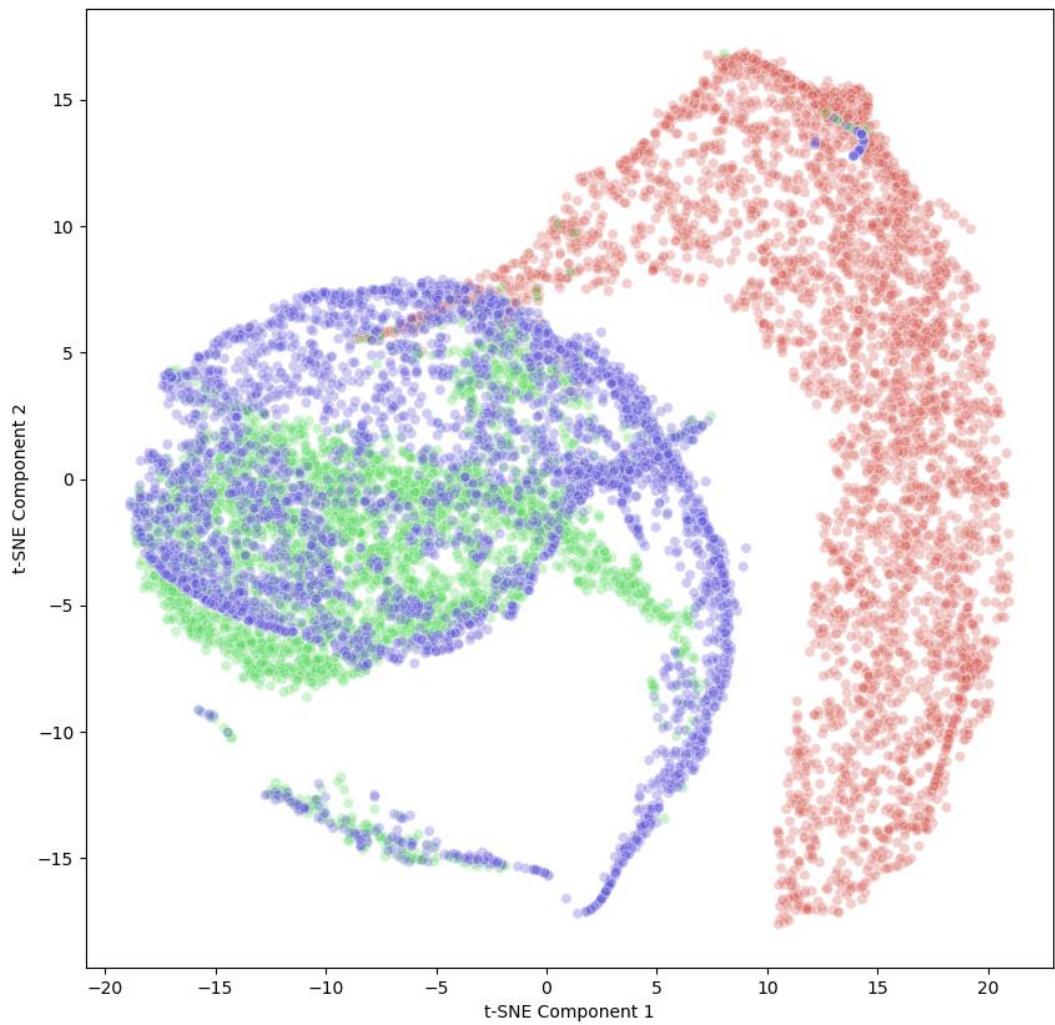


Figure 3.4. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 75; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

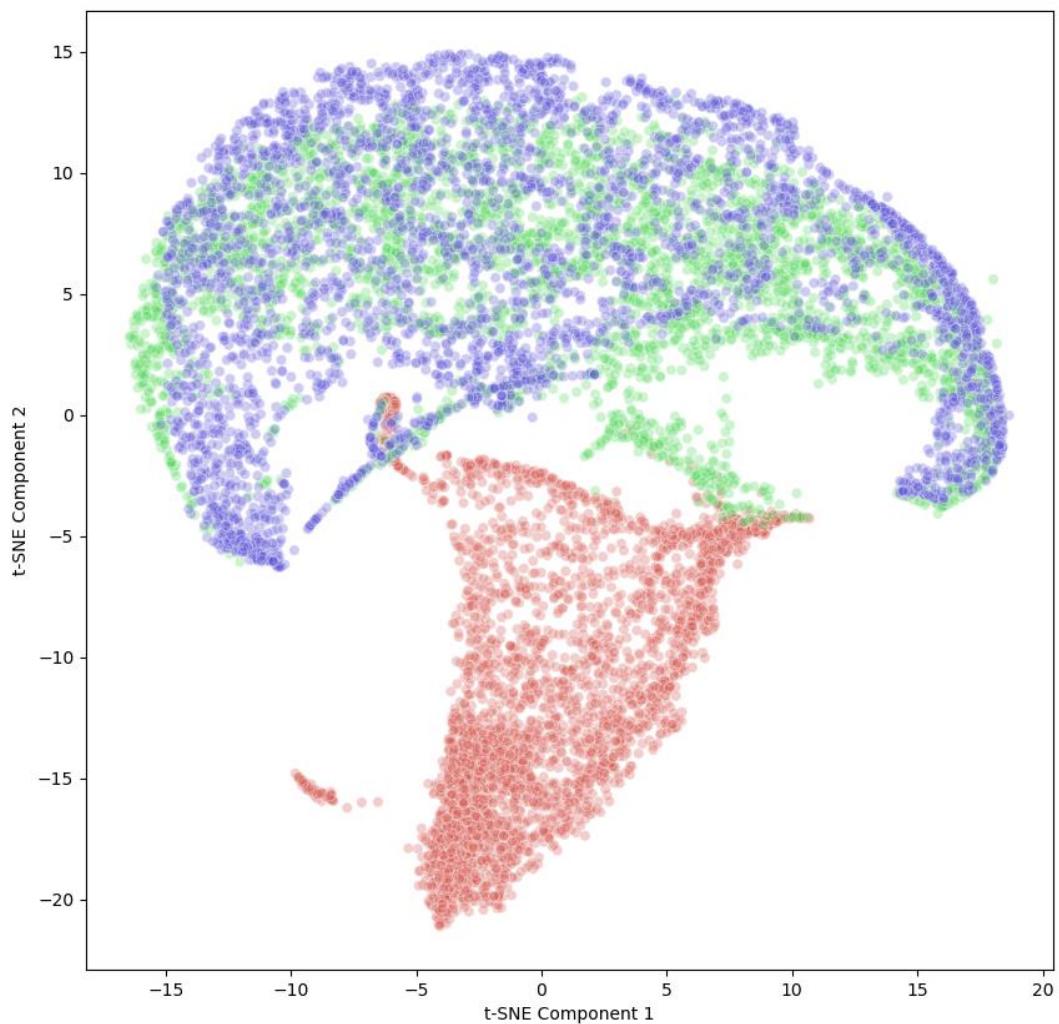


Figure 3.5. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **110**; early exaggeration = **12**; iterations = **1000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

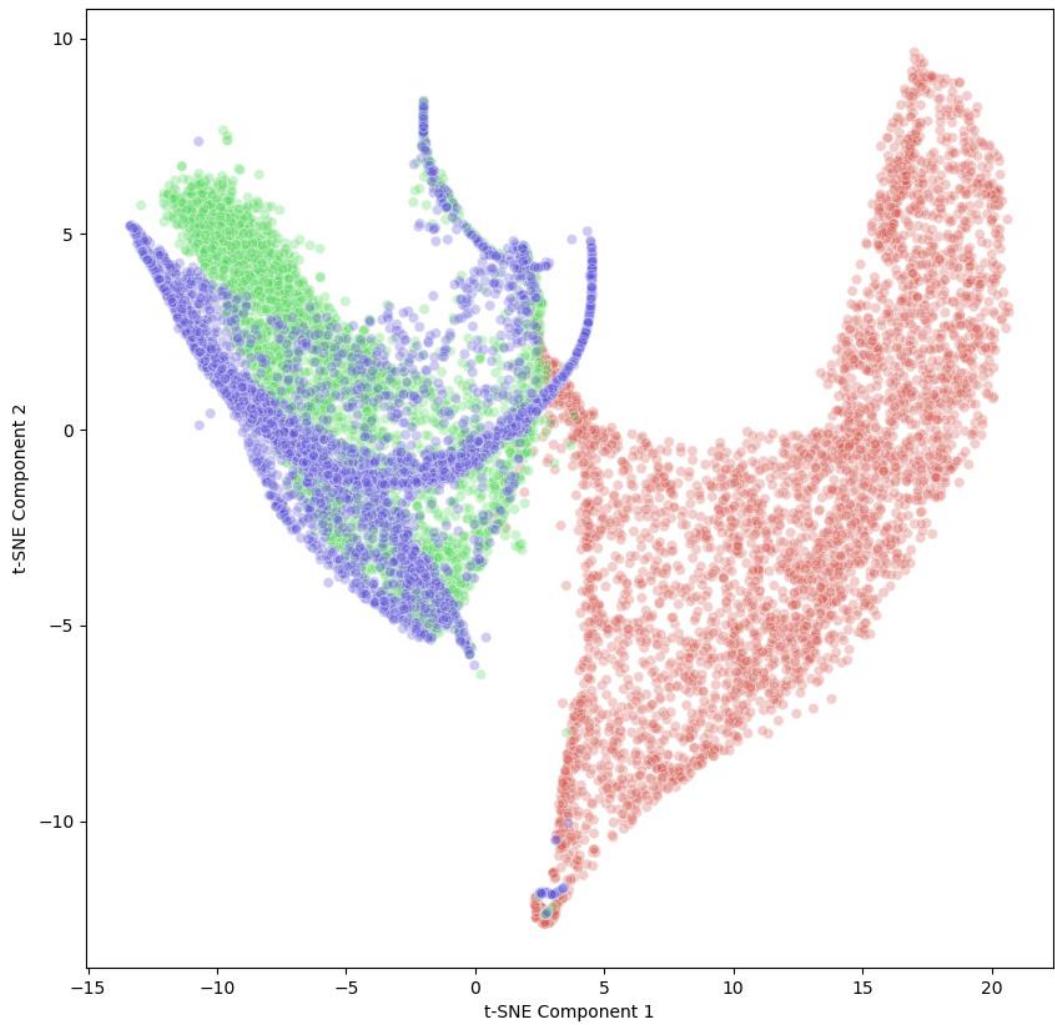


Figure 3.6. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 200; early exaggeration = 12; iterations = 1000. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

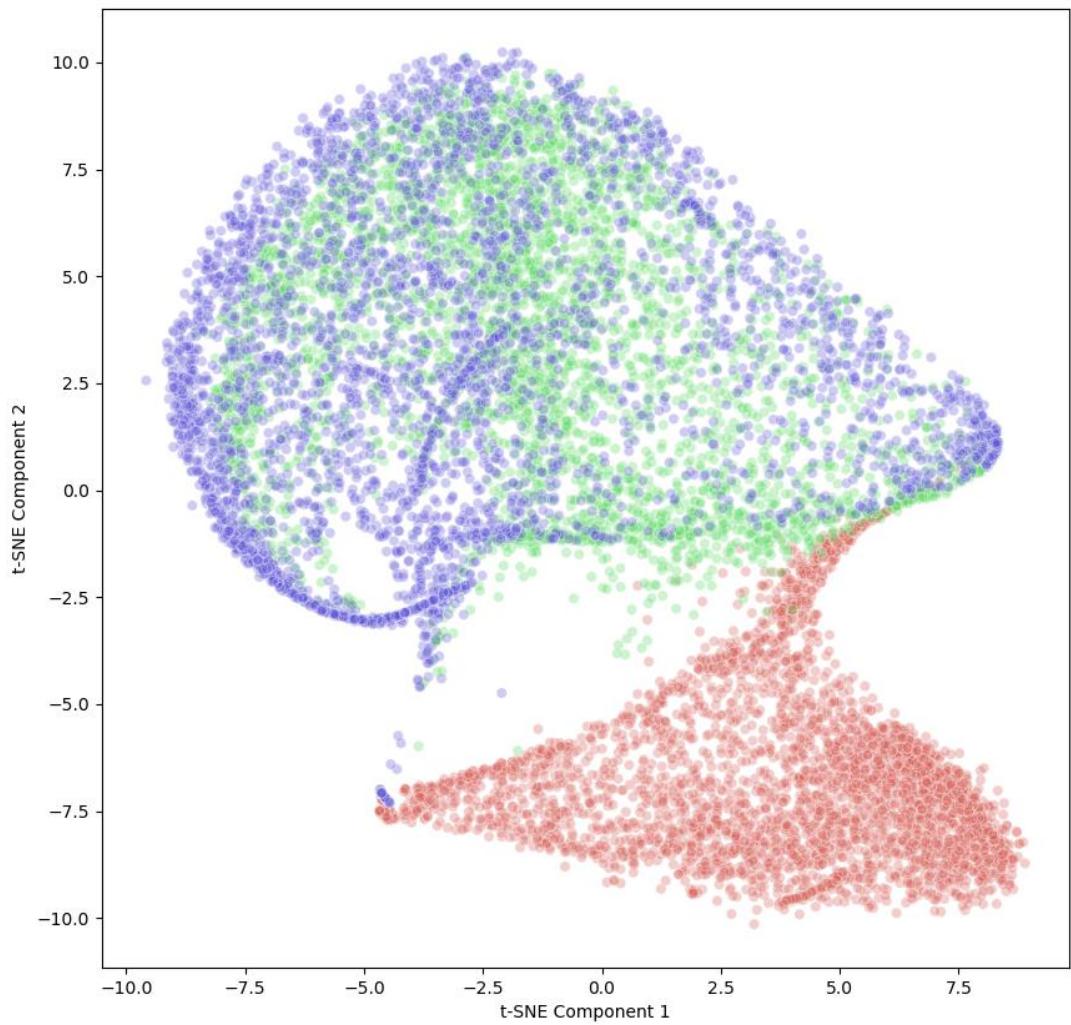


Figure 3.7. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **12**; iterations = **1000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

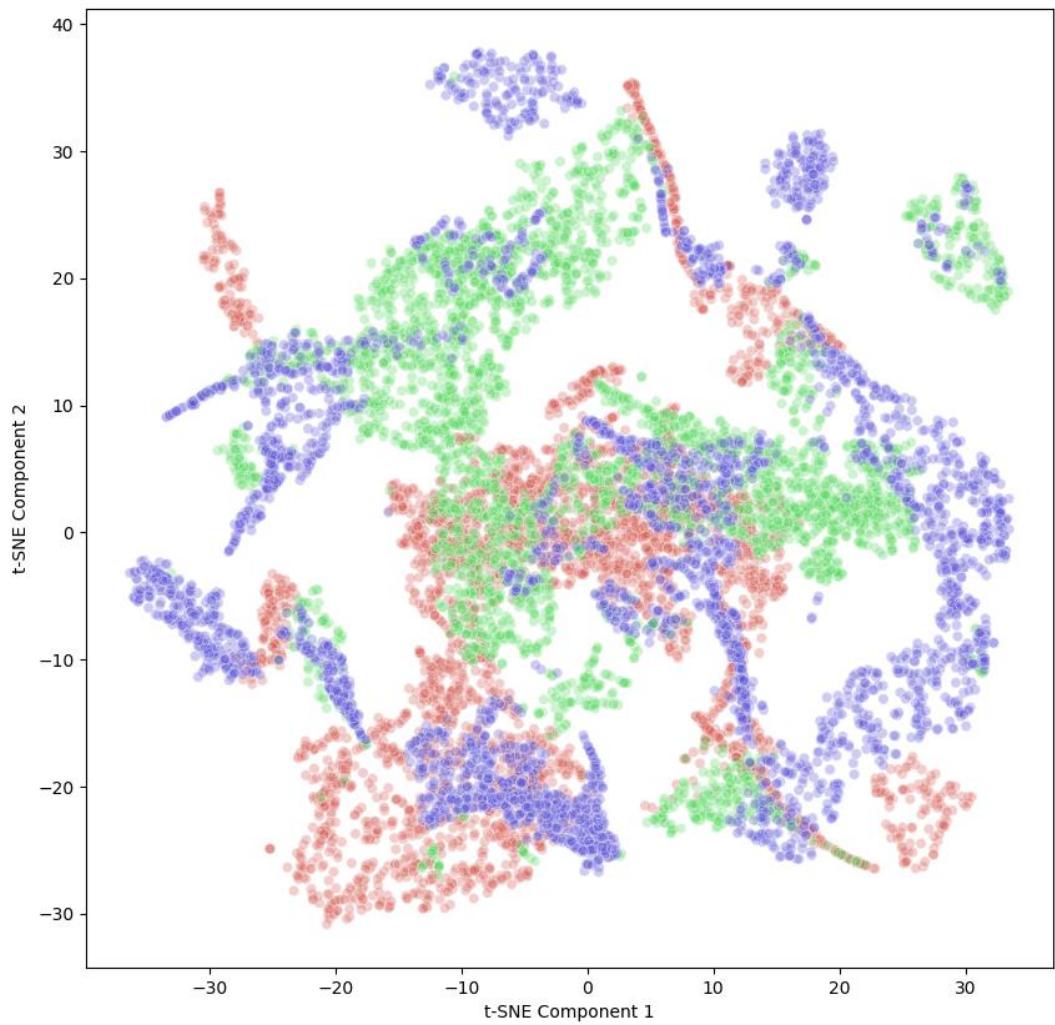


Figure 3.8. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 1; iterations = 1000. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

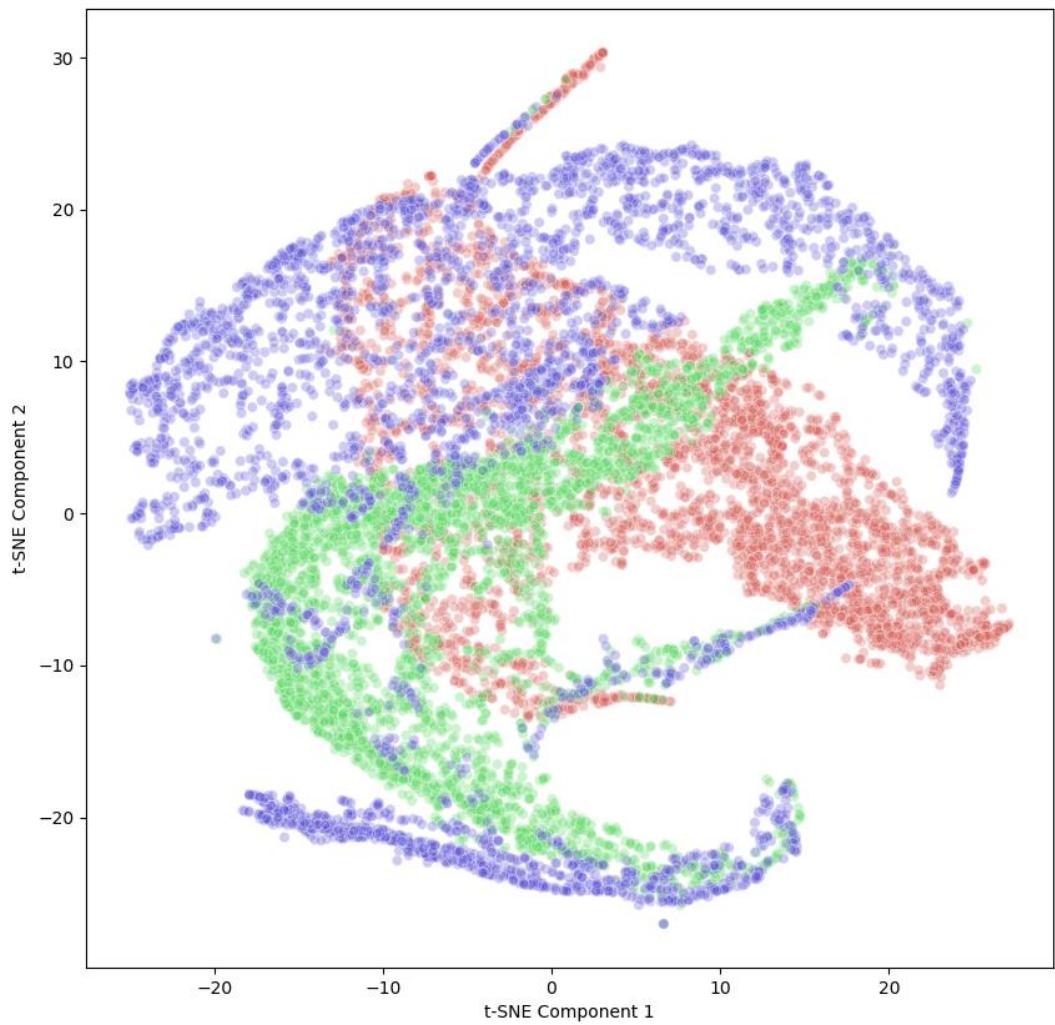


Figure 3.9. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 48; iterations = 1000. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

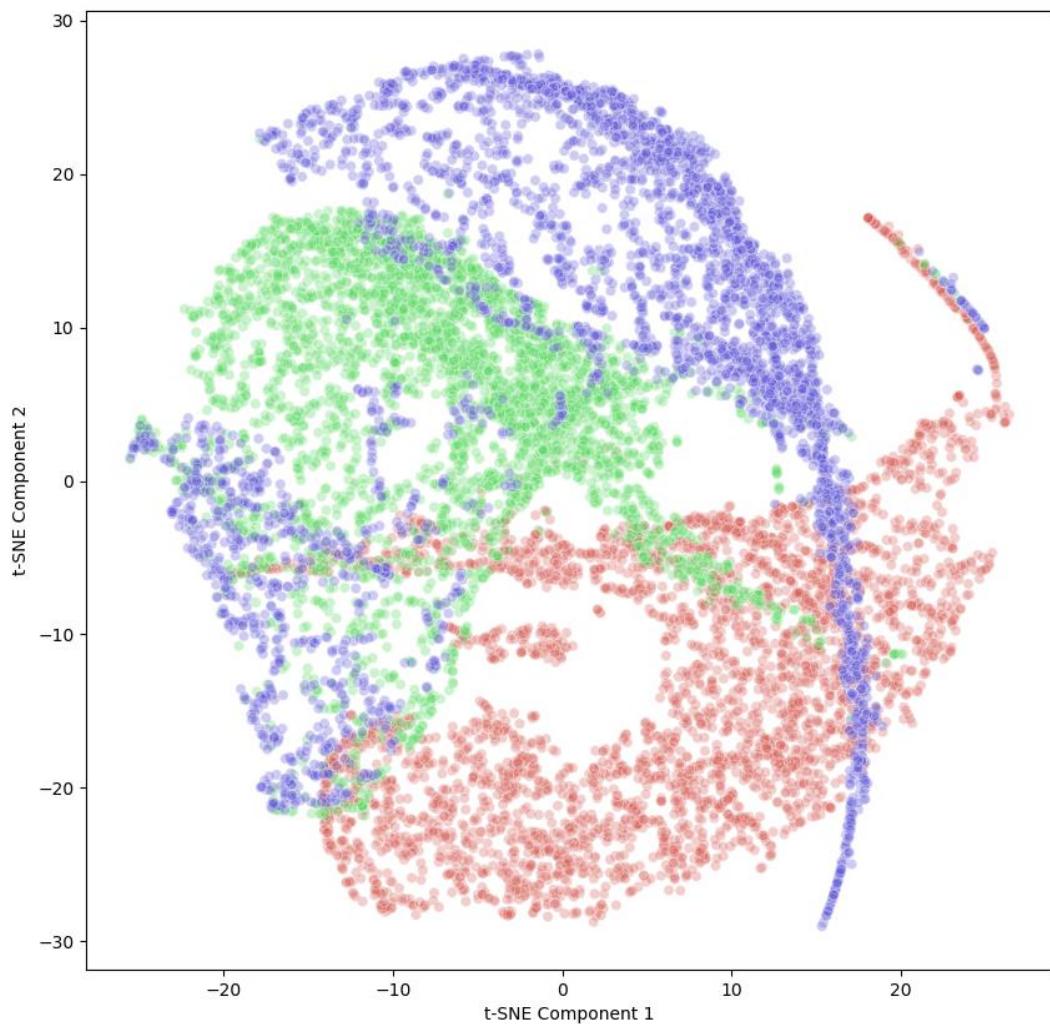


Figure 3.10. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 60; iterations = 1000. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

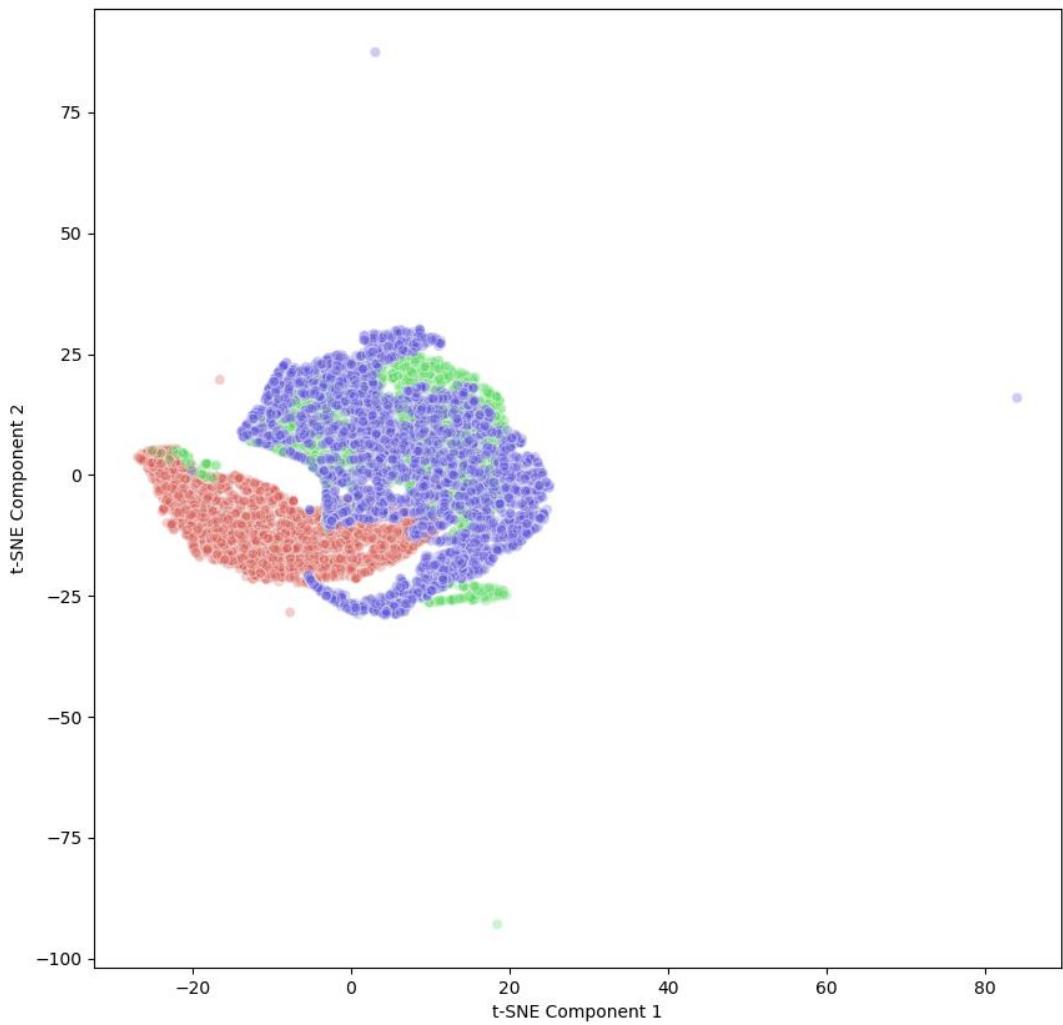


Figure 3.11. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 72; iterations = 1000. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

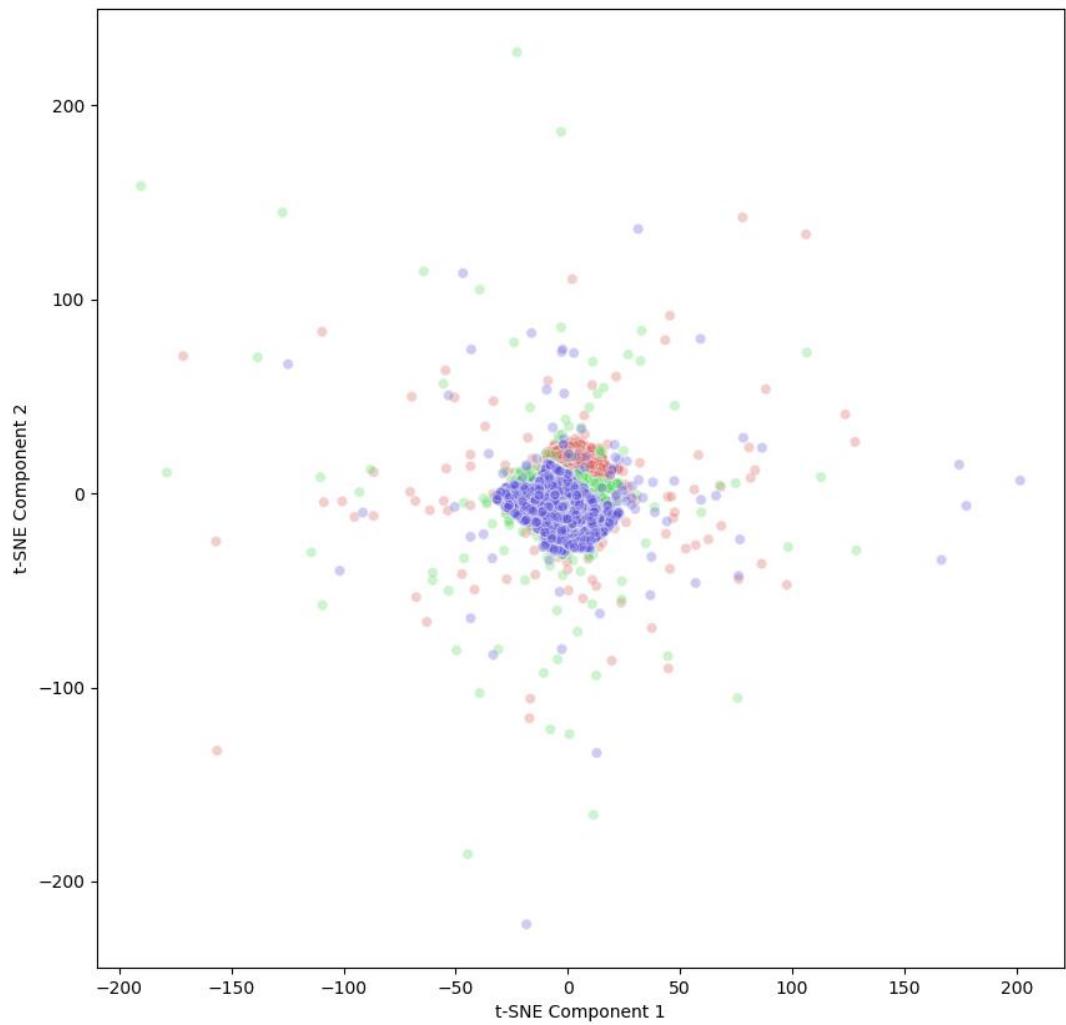


Figure 3.12. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 96; iterations = 1000. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

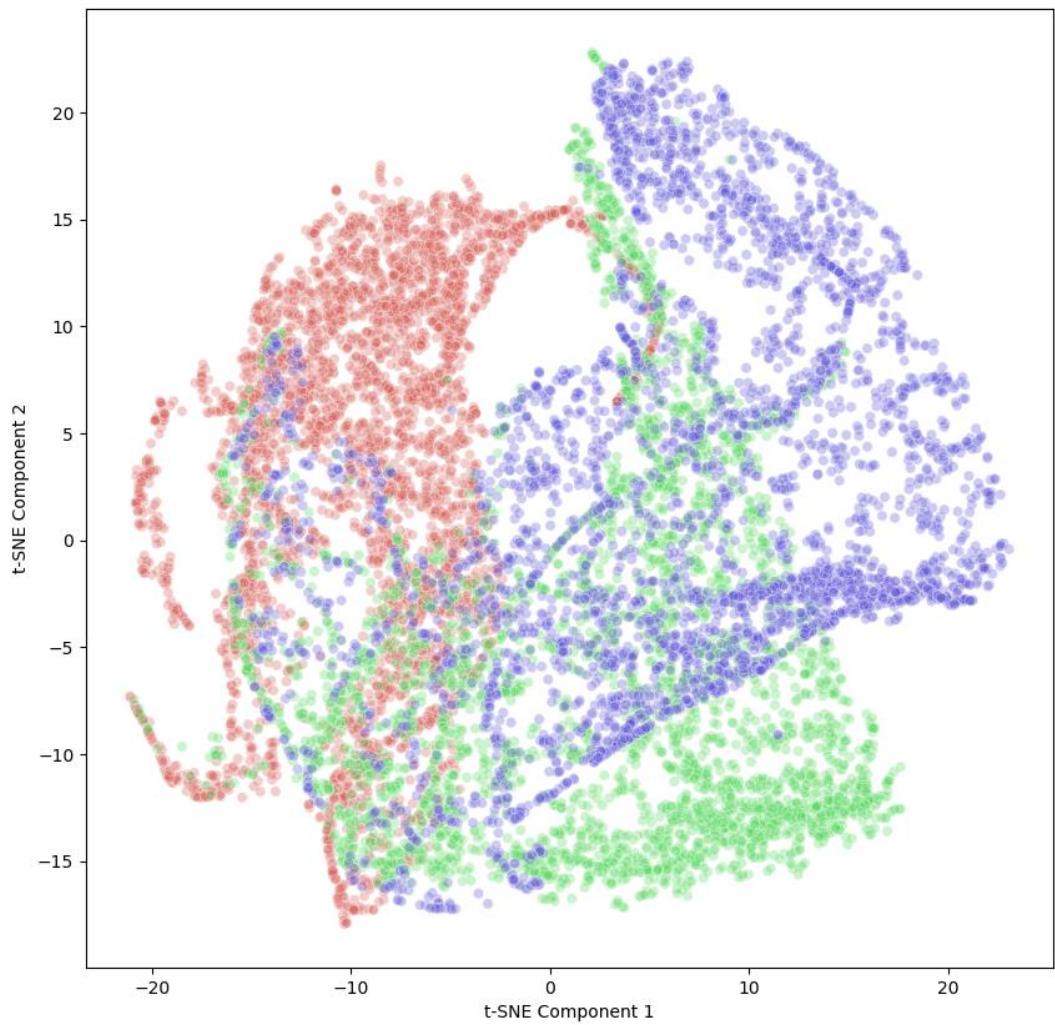


Figure 3.13. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 12; iterations = 500. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

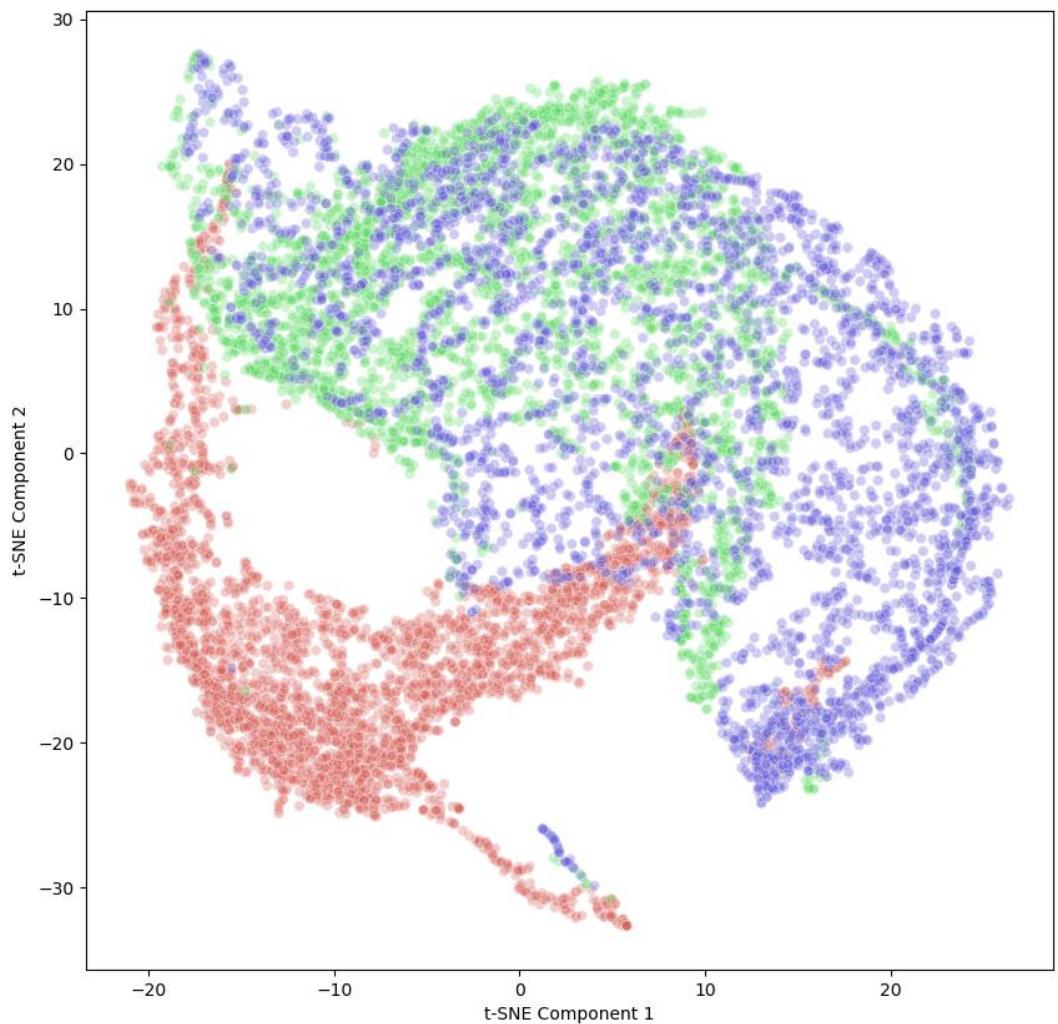


Figure 3.14. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = 30; early exaggeration = 12; iterations = 2000. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

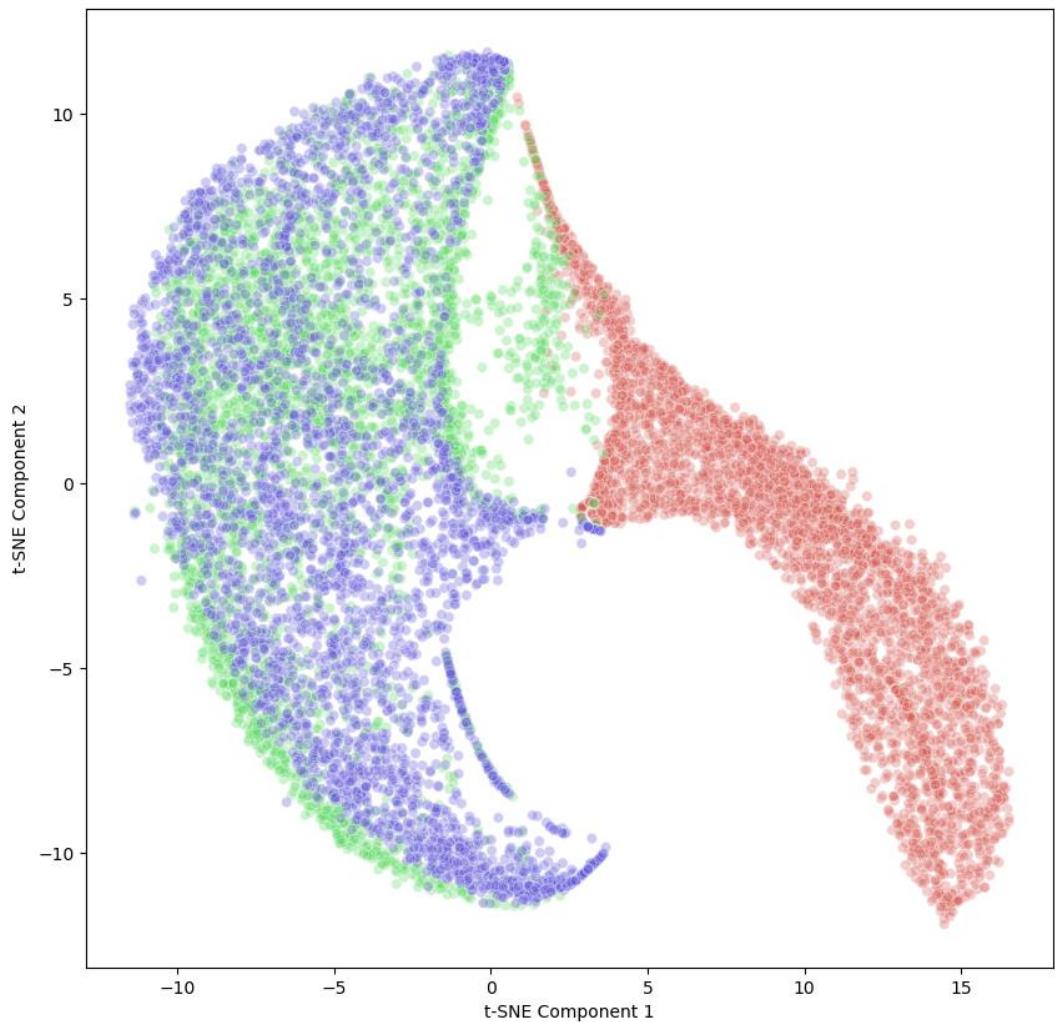


Figure 3.15. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **48**; iterations = **2000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

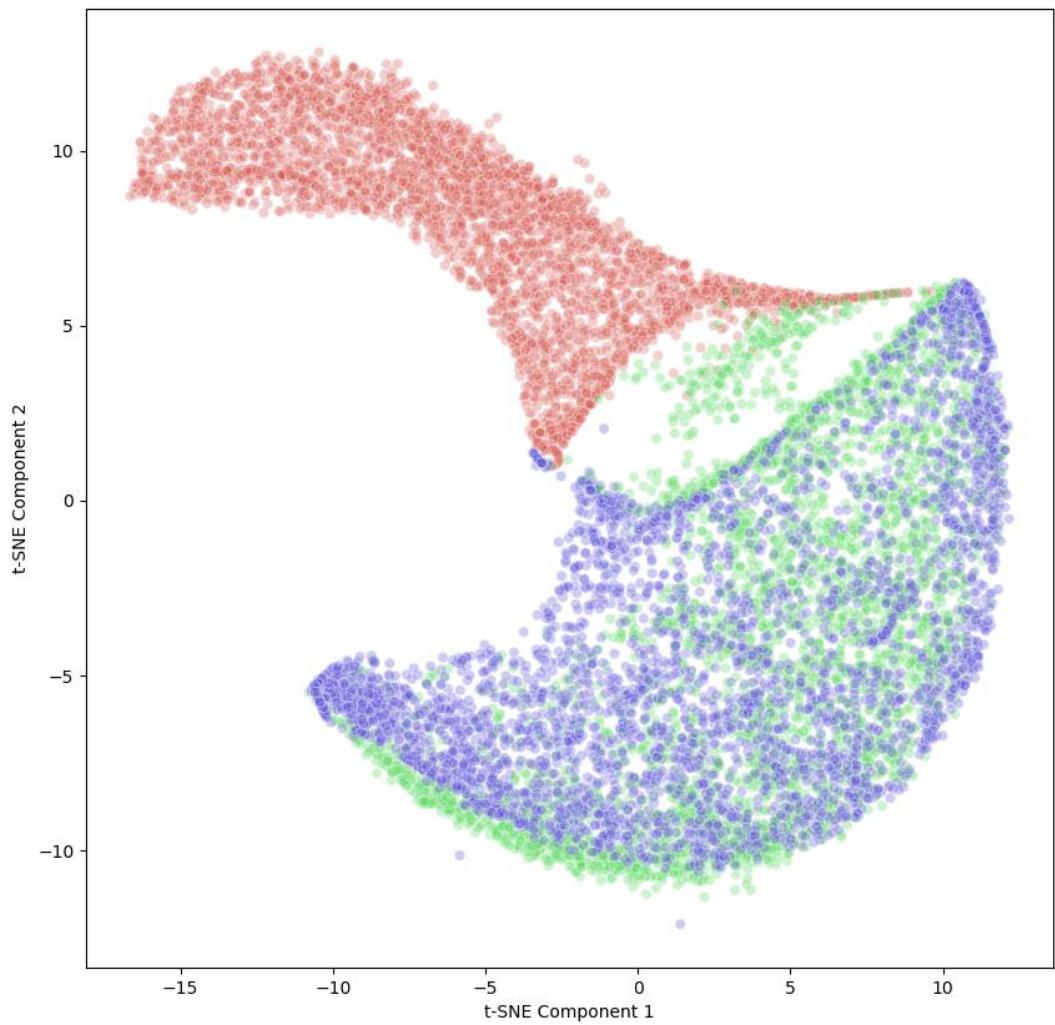


Figure 3.16. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **48**; iterations = **1000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

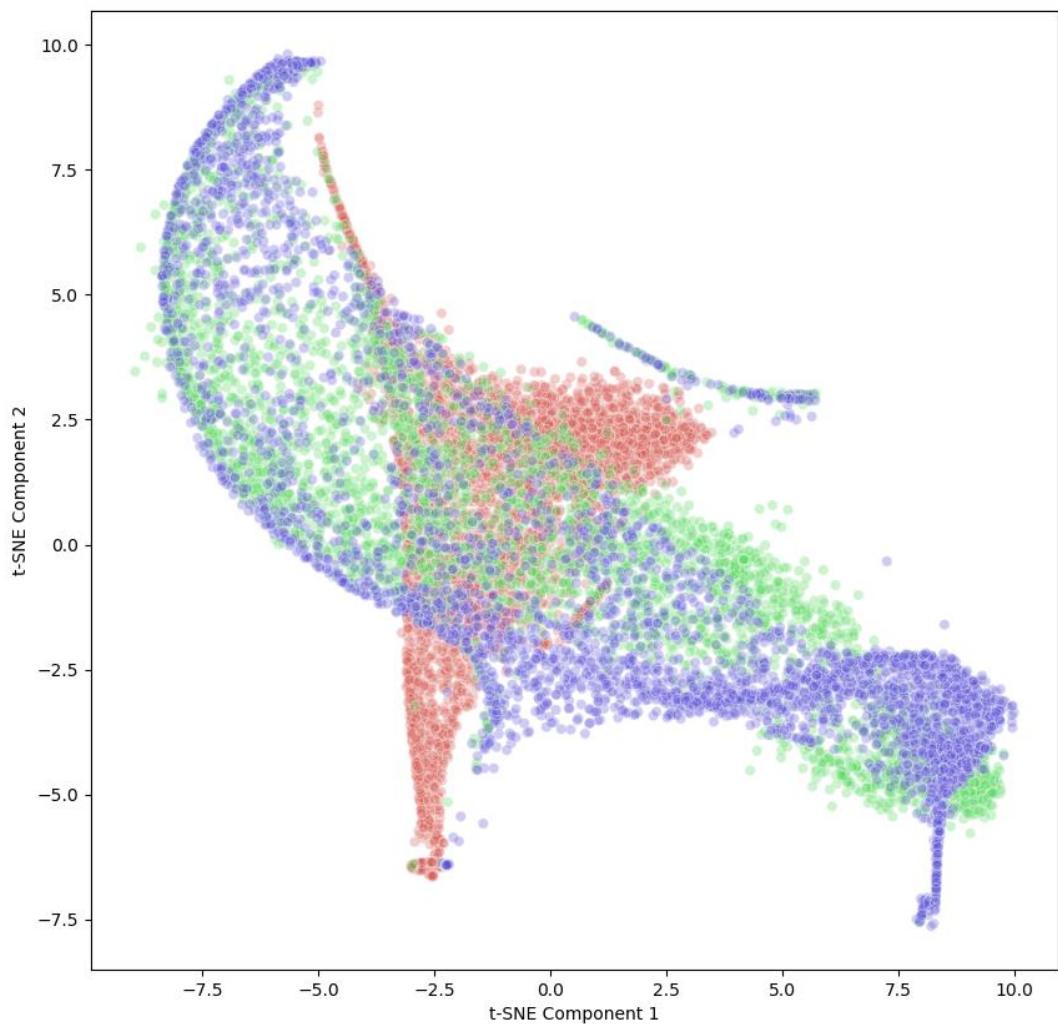


Figure 3.17. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 500, perplexity = **400**; early exaggeration = **60**; iterations = **2000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

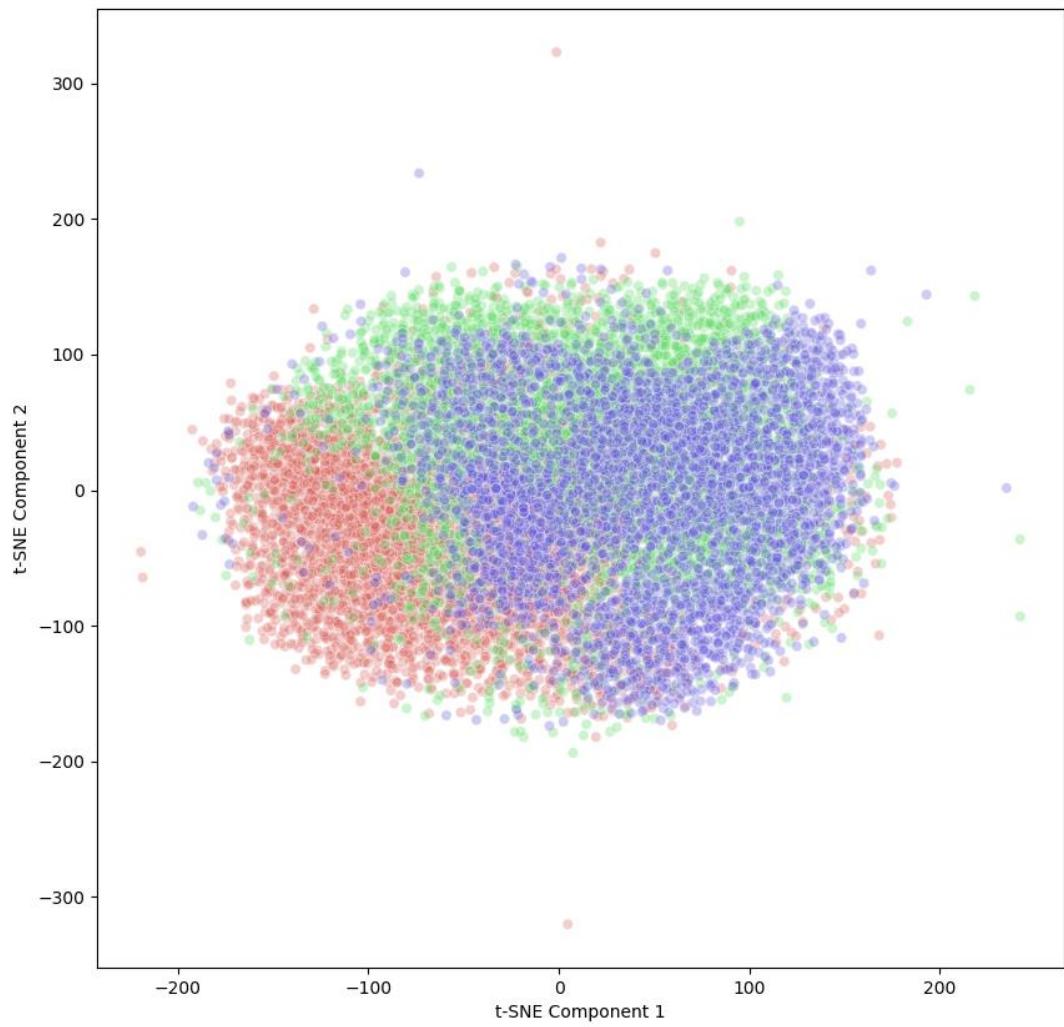


Figure 3.18. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **1000**, perplexity = **400**; early exaggeration = **60**; iterations = **2000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

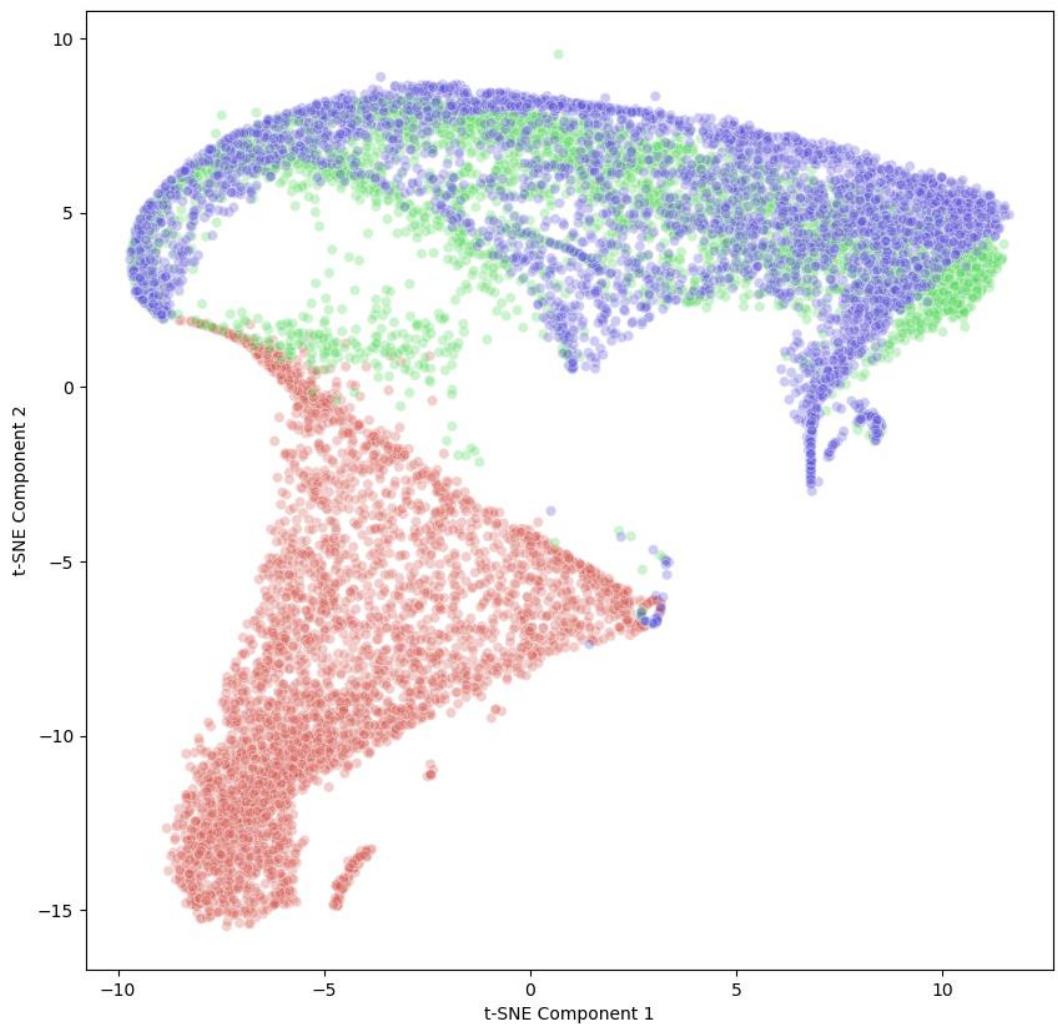


Figure 3.19. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **10**, perplexity = **400**; early exaggeration = **60**; iterations = **5000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

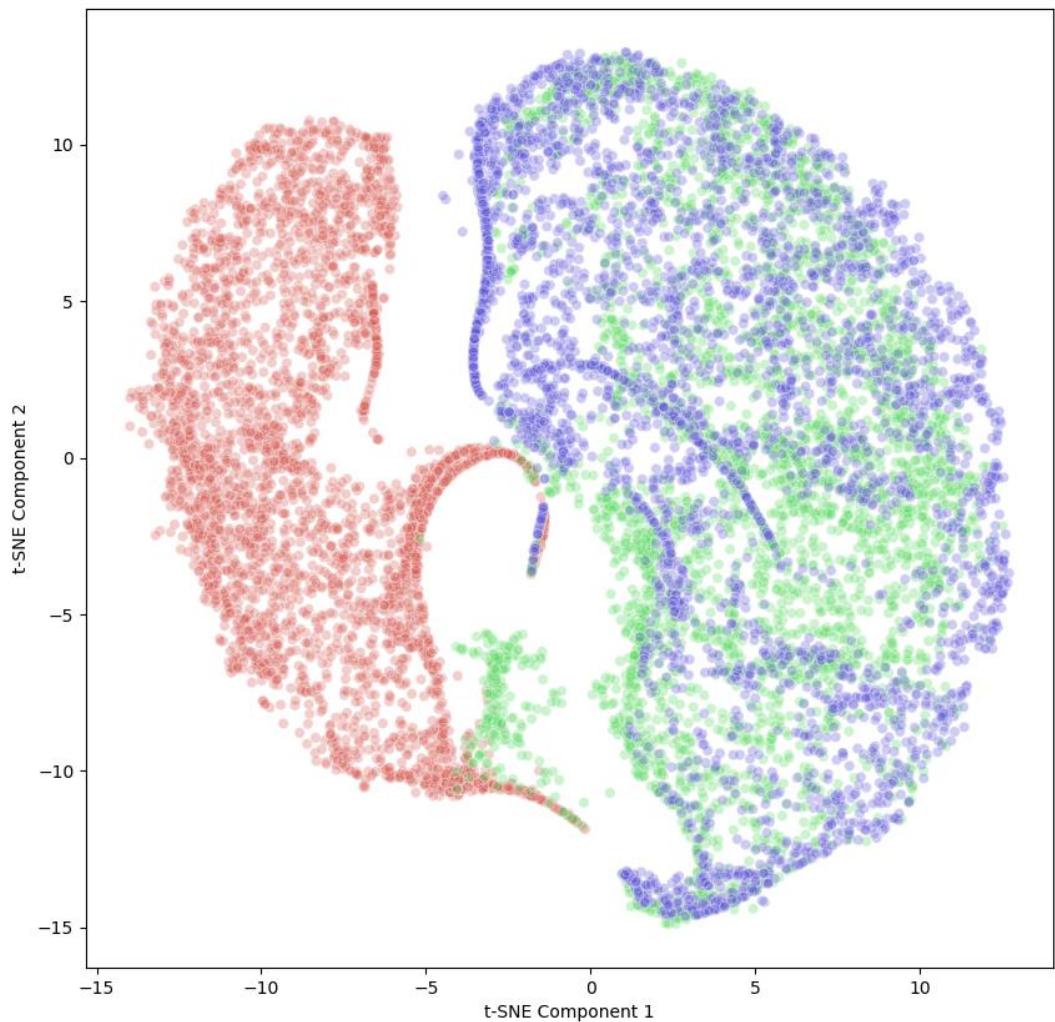


Figure 3.20. t-SNE visualisation: components = 3; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **10**, perplexity = **110**; early exaggeration = **60**; iterations = **1000**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

#### 4 Uniform Manifold Approximation and Projection, *UMAP*

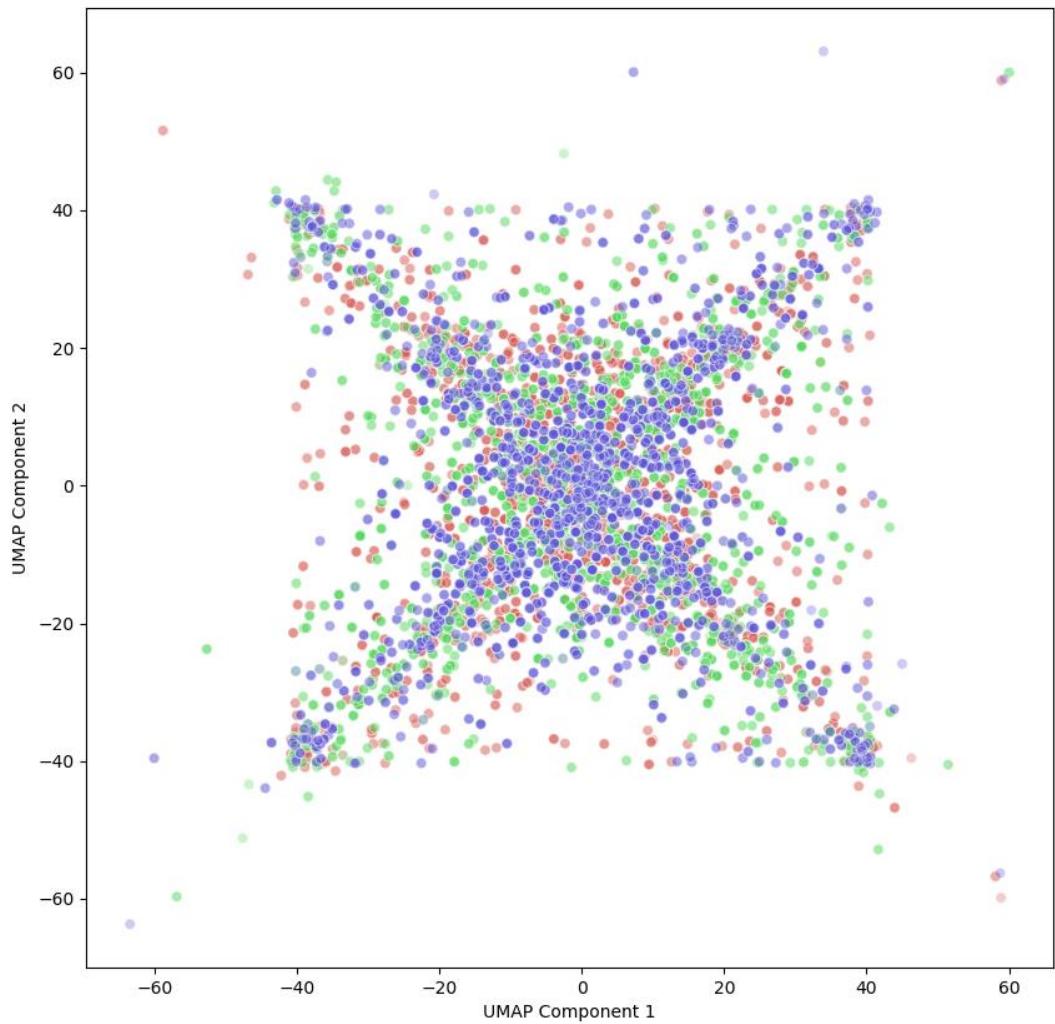


Figure 4.1. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 2; minimum distance = **0.1**; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

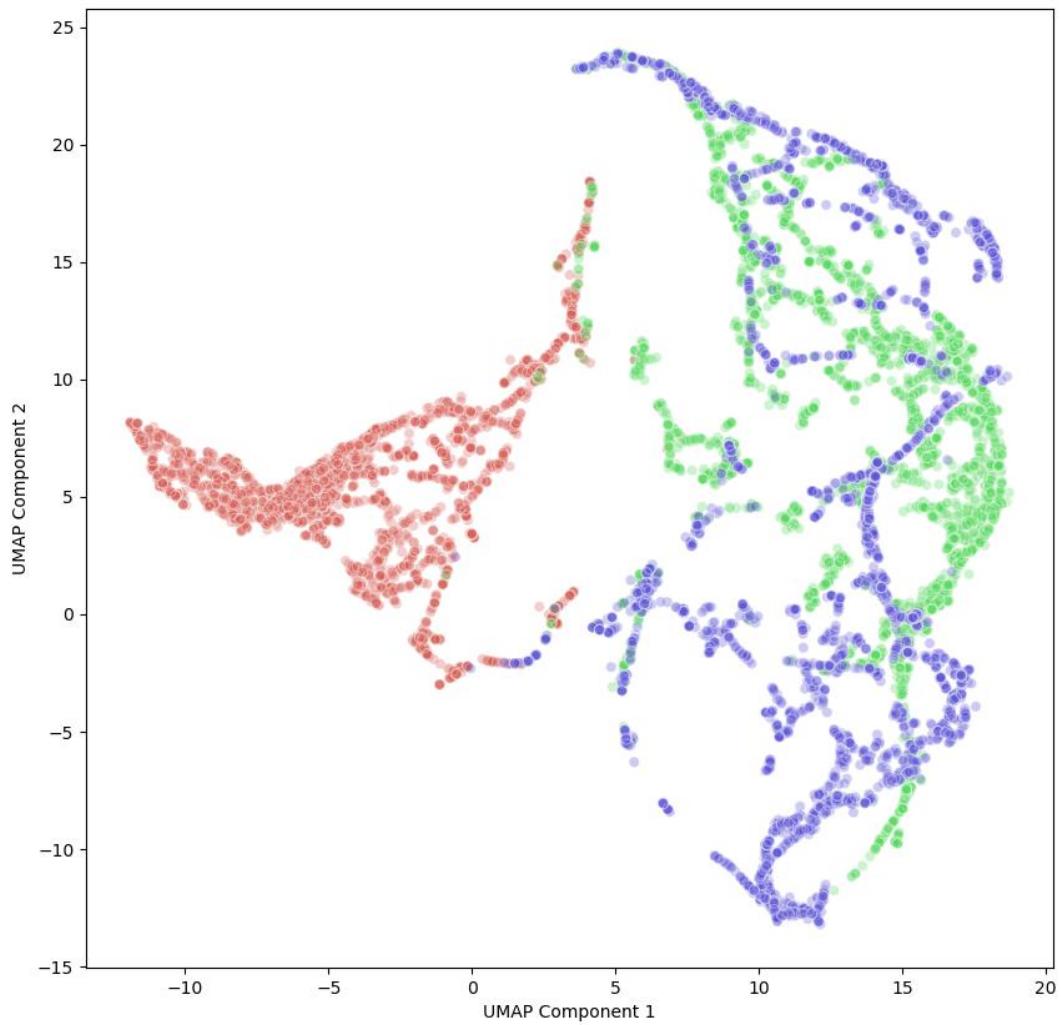


Figure 4.2. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 5; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

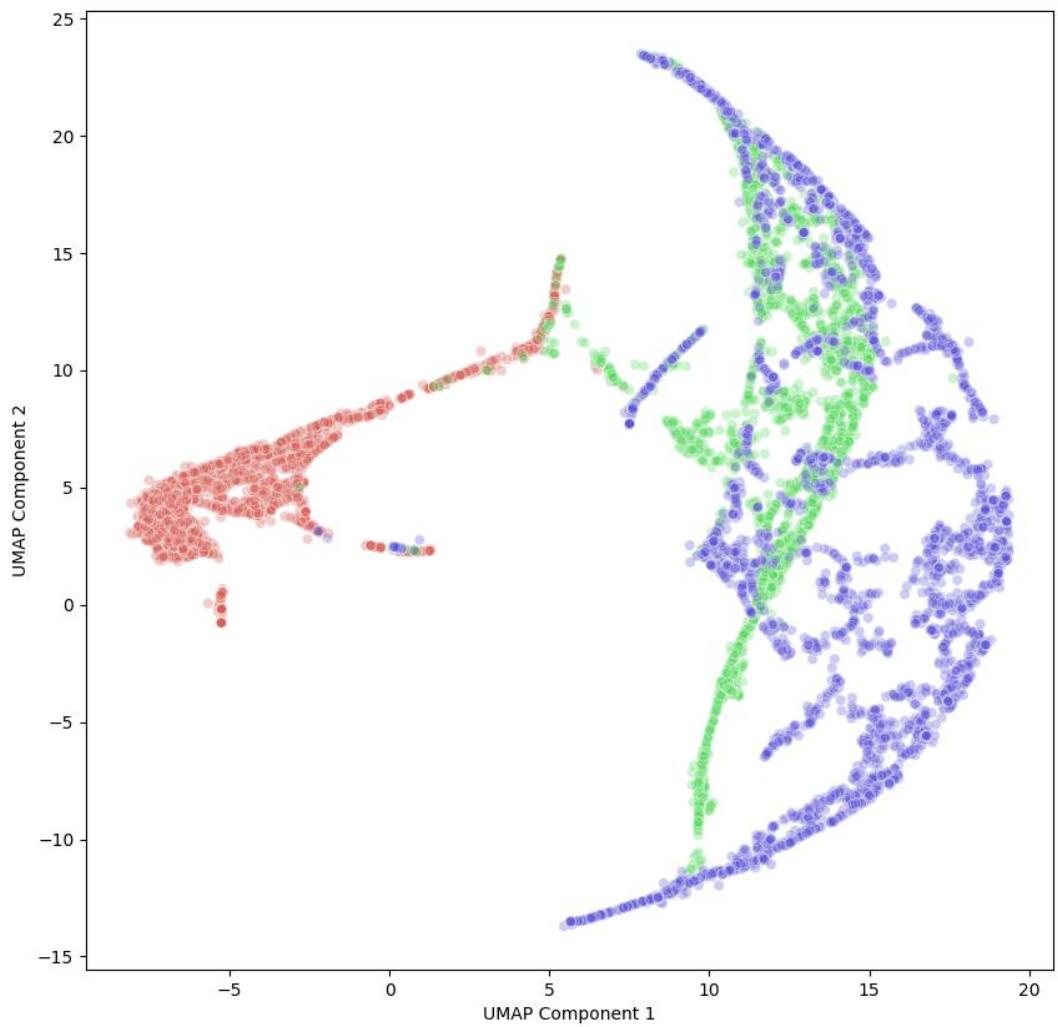


Figure 4.3. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

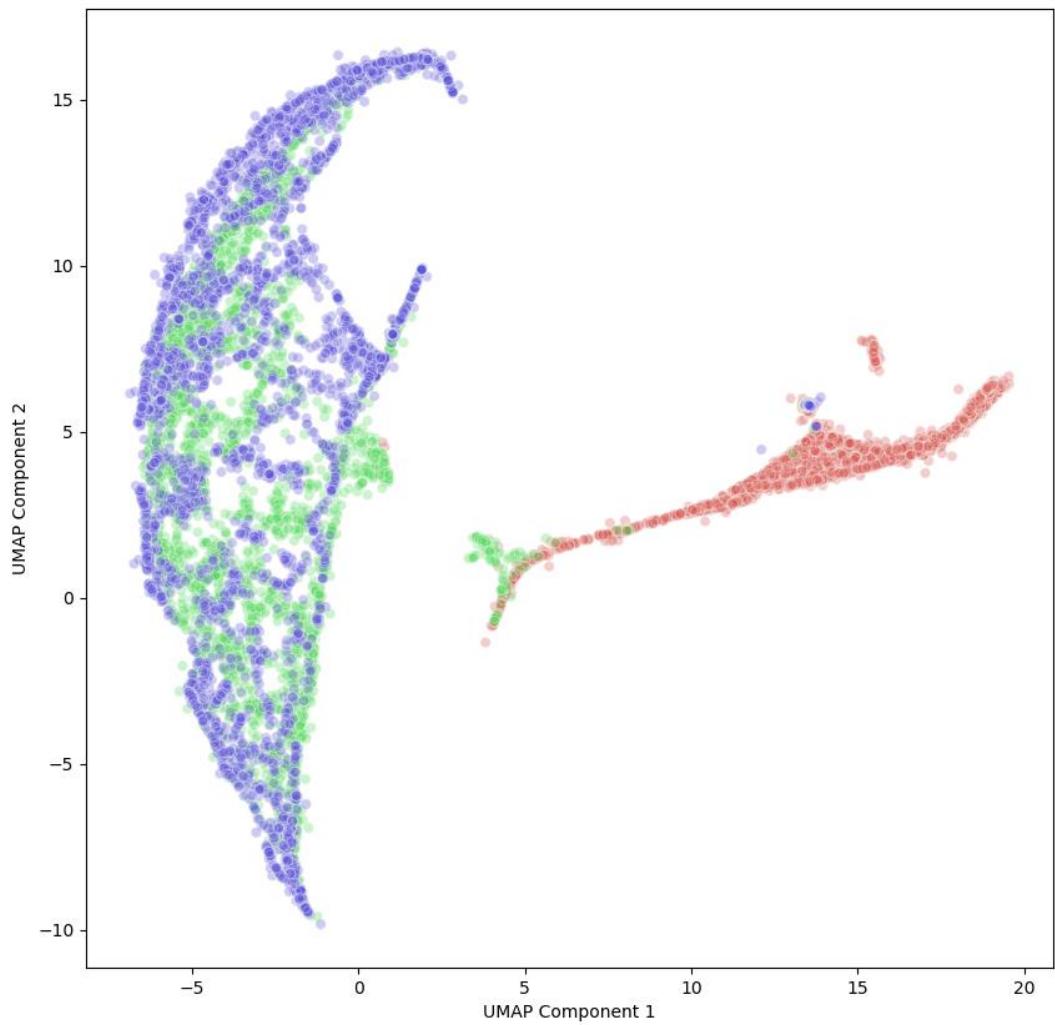


Figure 4.4. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 50; minimum distance = 0.1; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

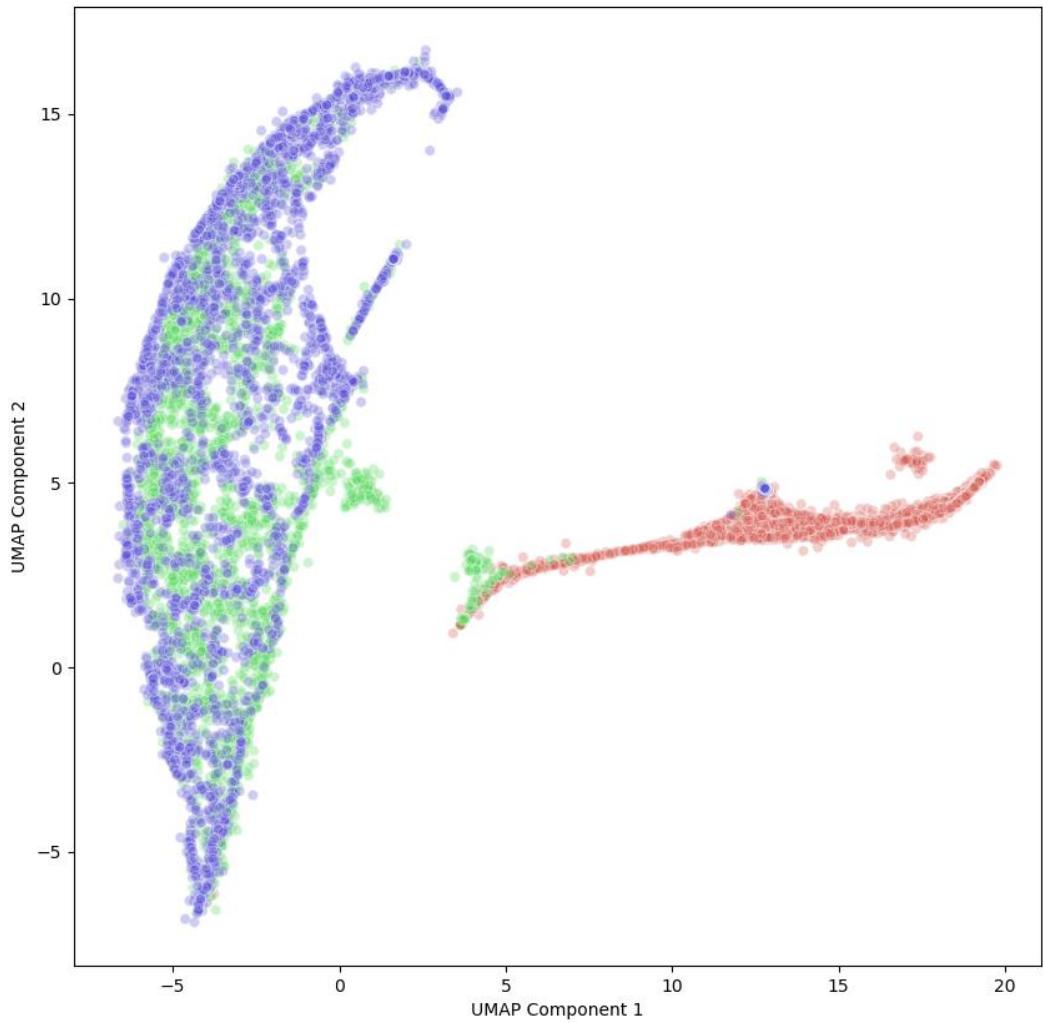


Figure 4.5. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **100**; minimum distance = **0.1**; transform queue size = **4**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

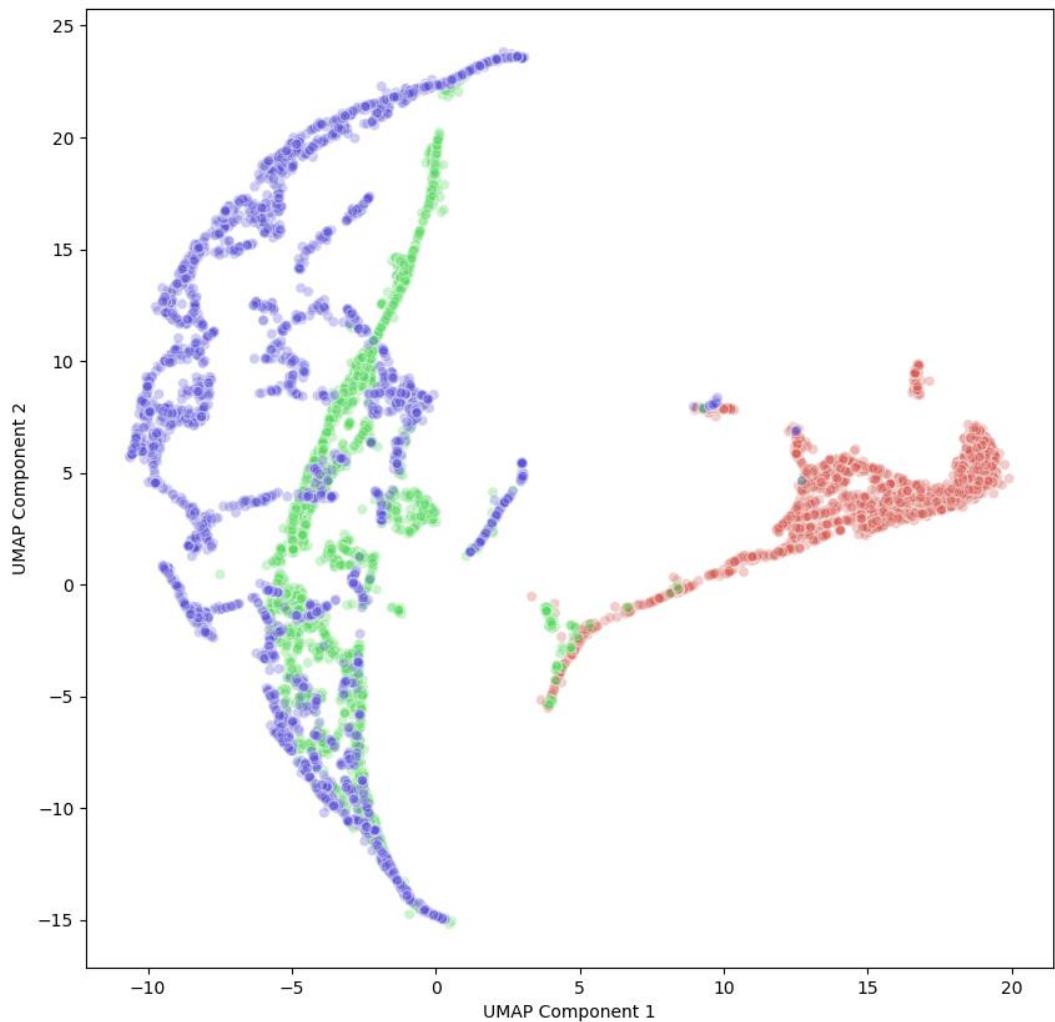


Figure 4.6. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.0; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

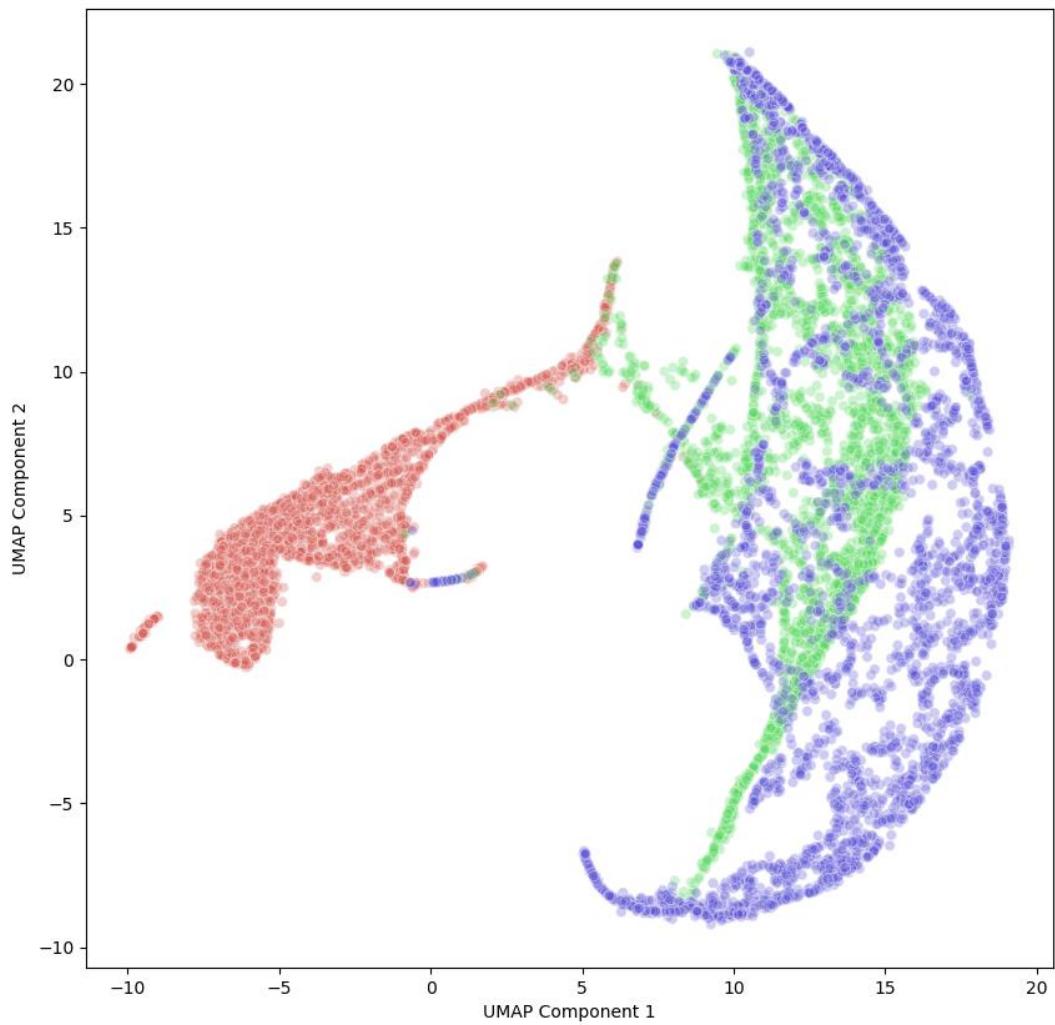


Figure 4.7. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = 0.25; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

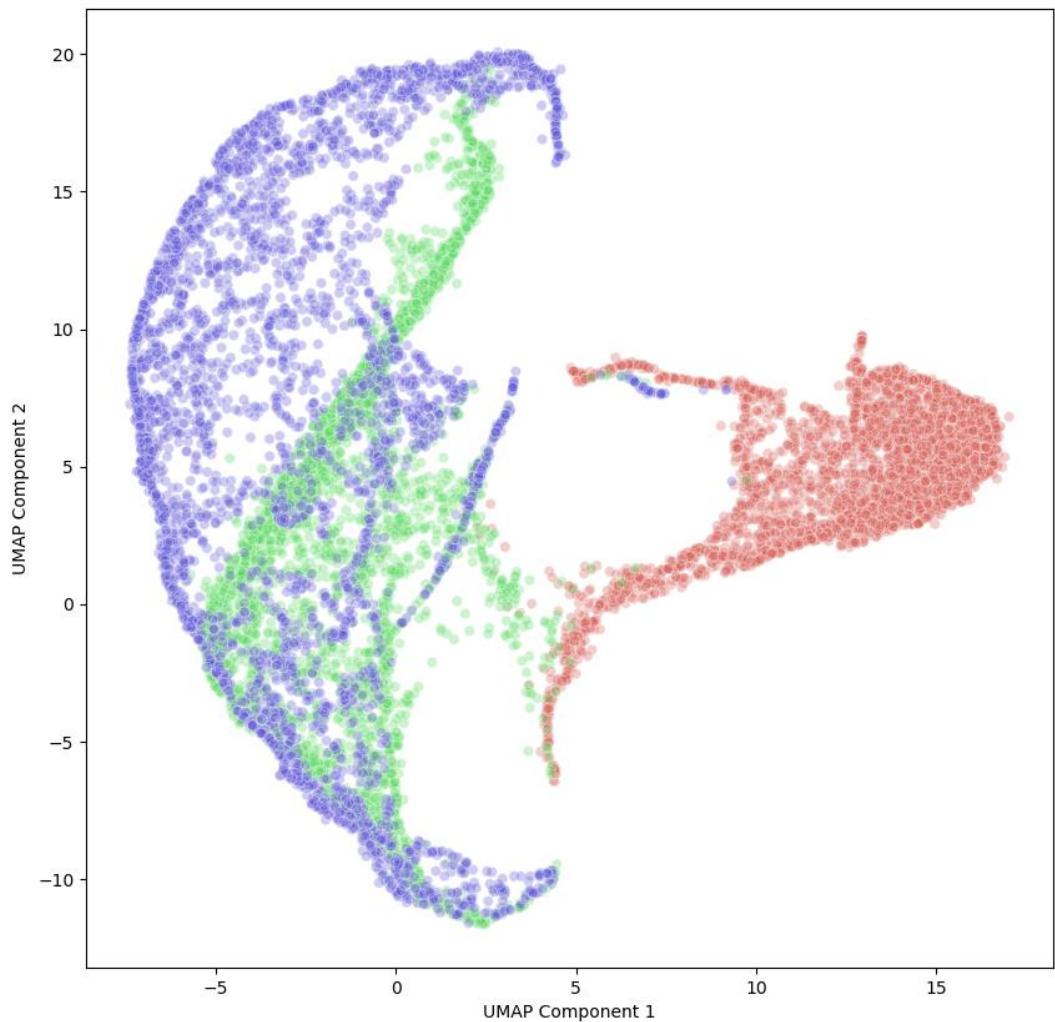


Figure 4.8. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.50; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

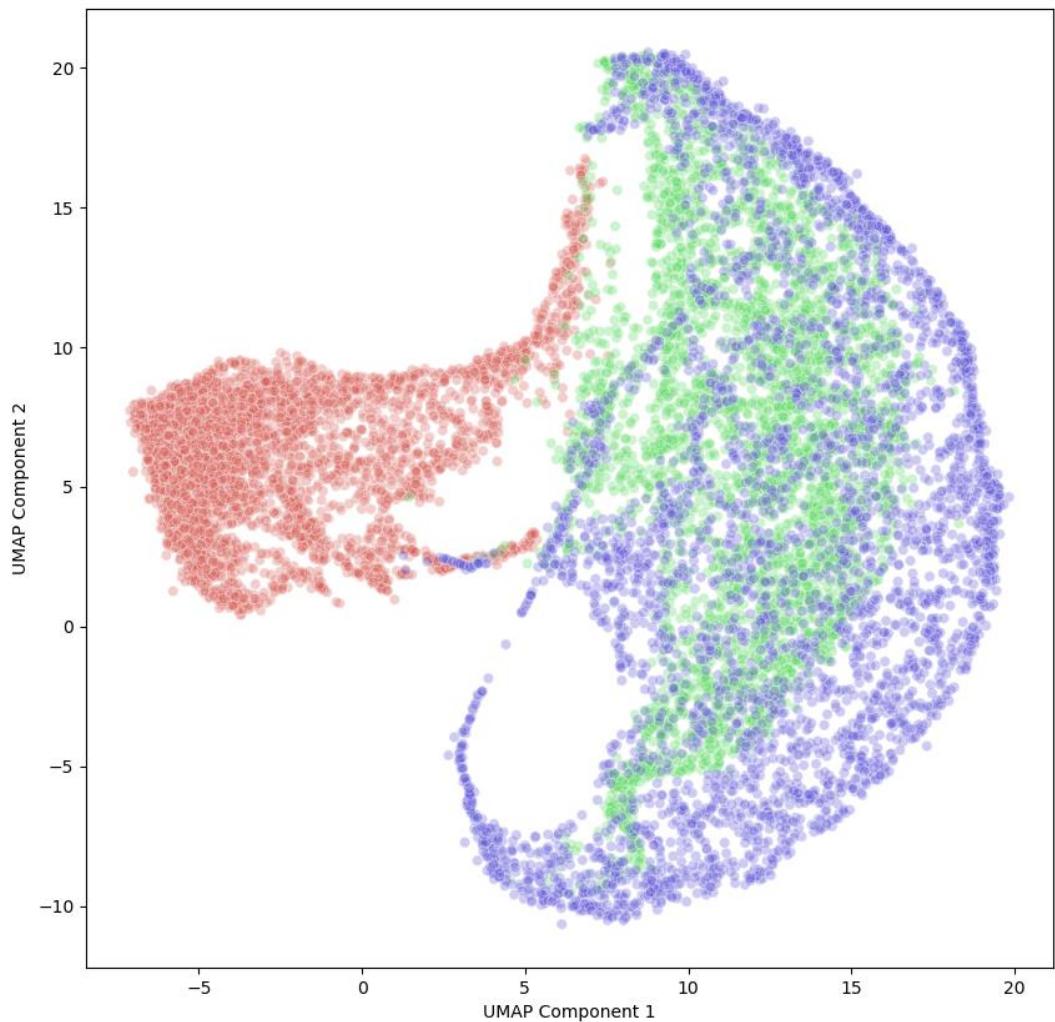


Figure 4.9. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.80; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

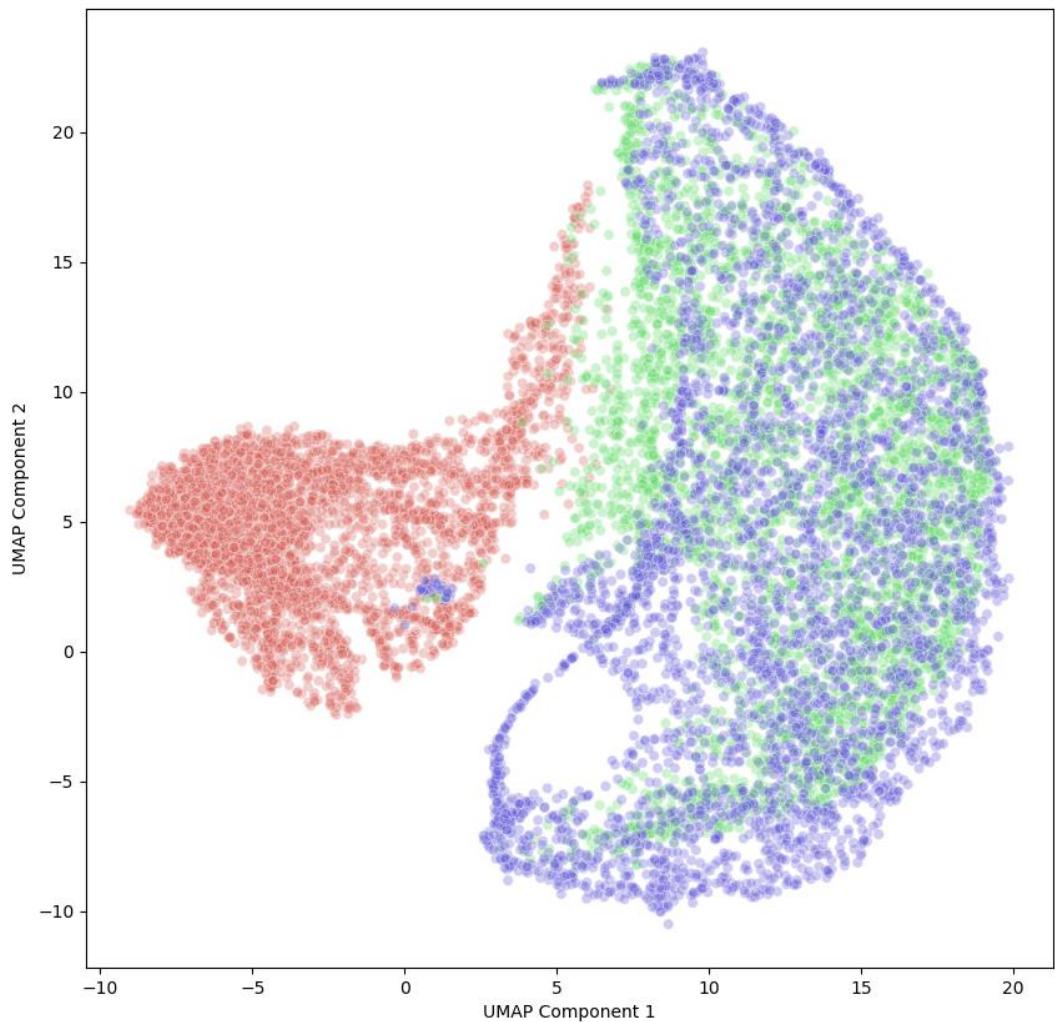


Figure 4.10. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 15; minimum distance = **0.99**; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

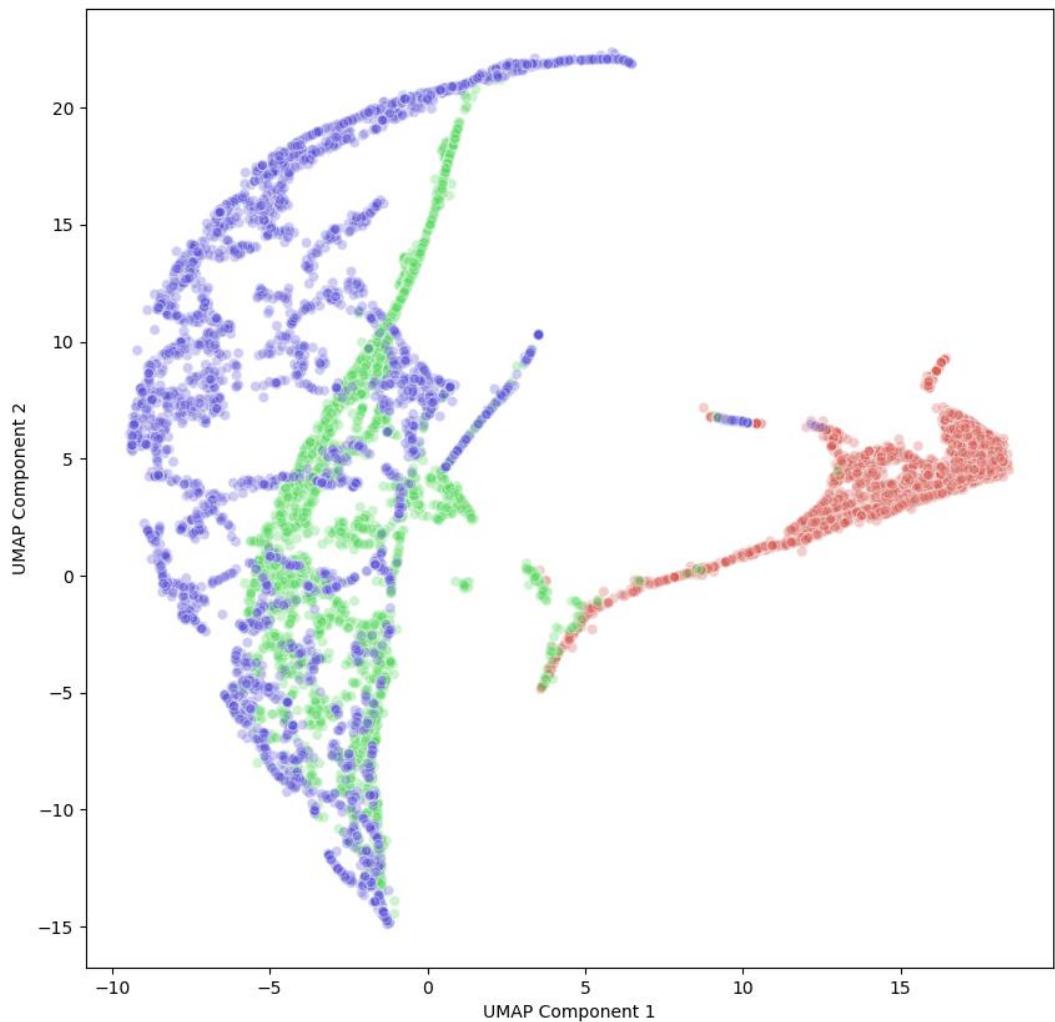


Figure 4.11. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 50. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

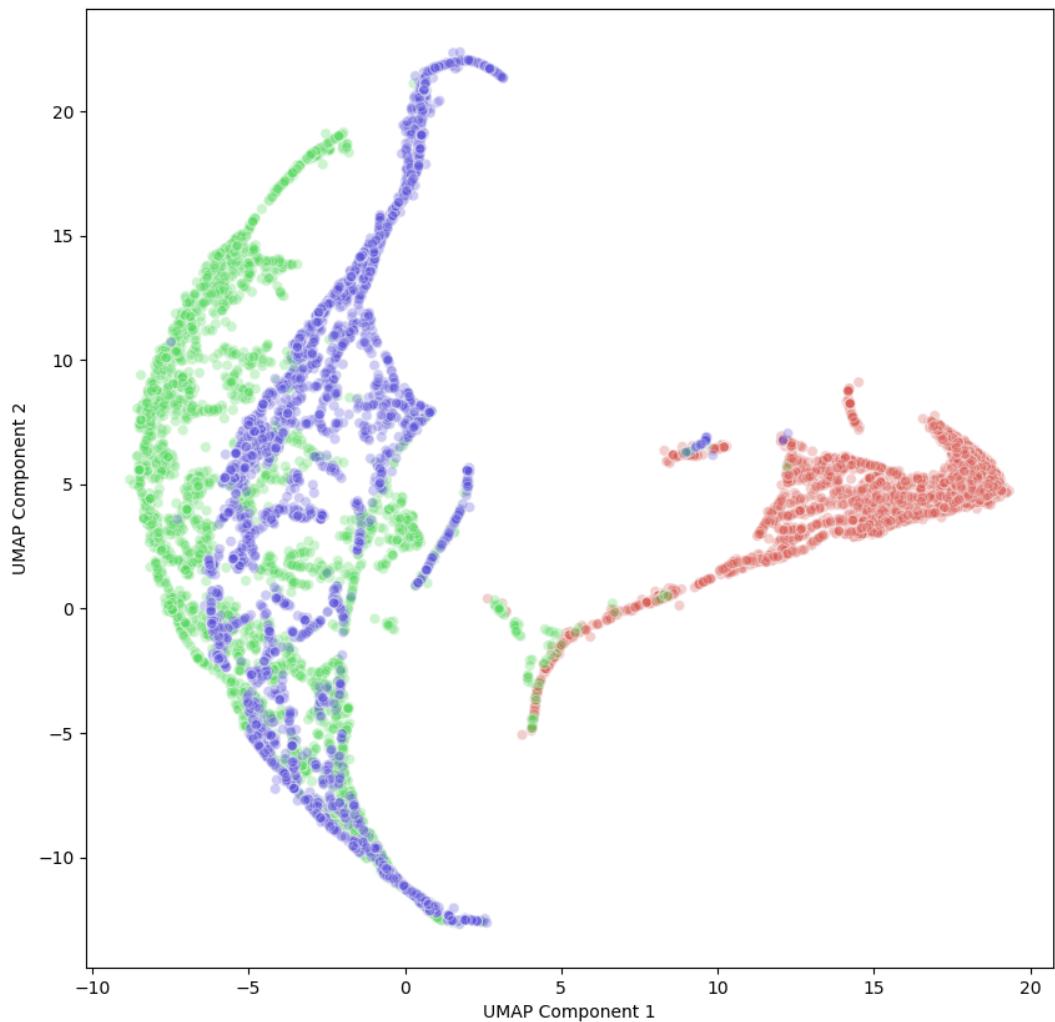


Figure 4.12. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 75. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

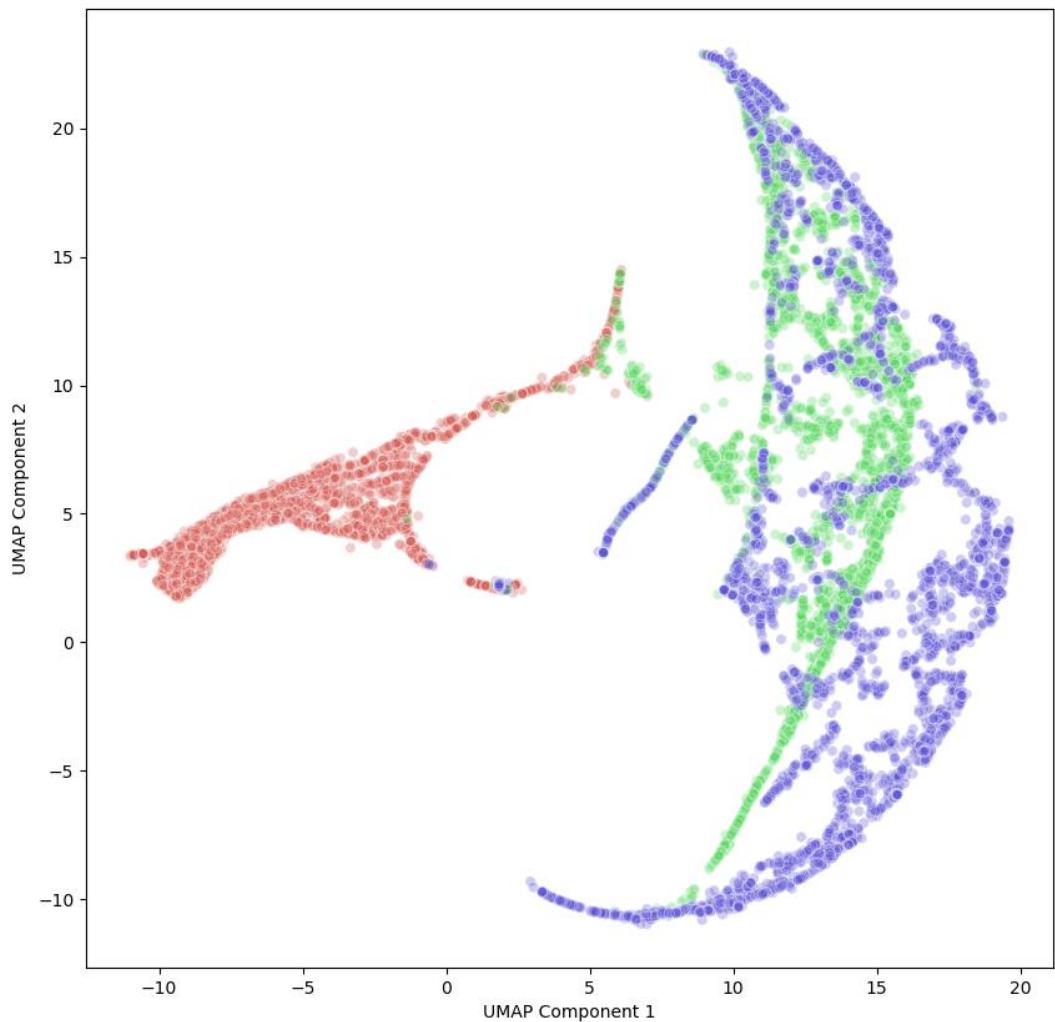


Figure 4.13. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 100. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

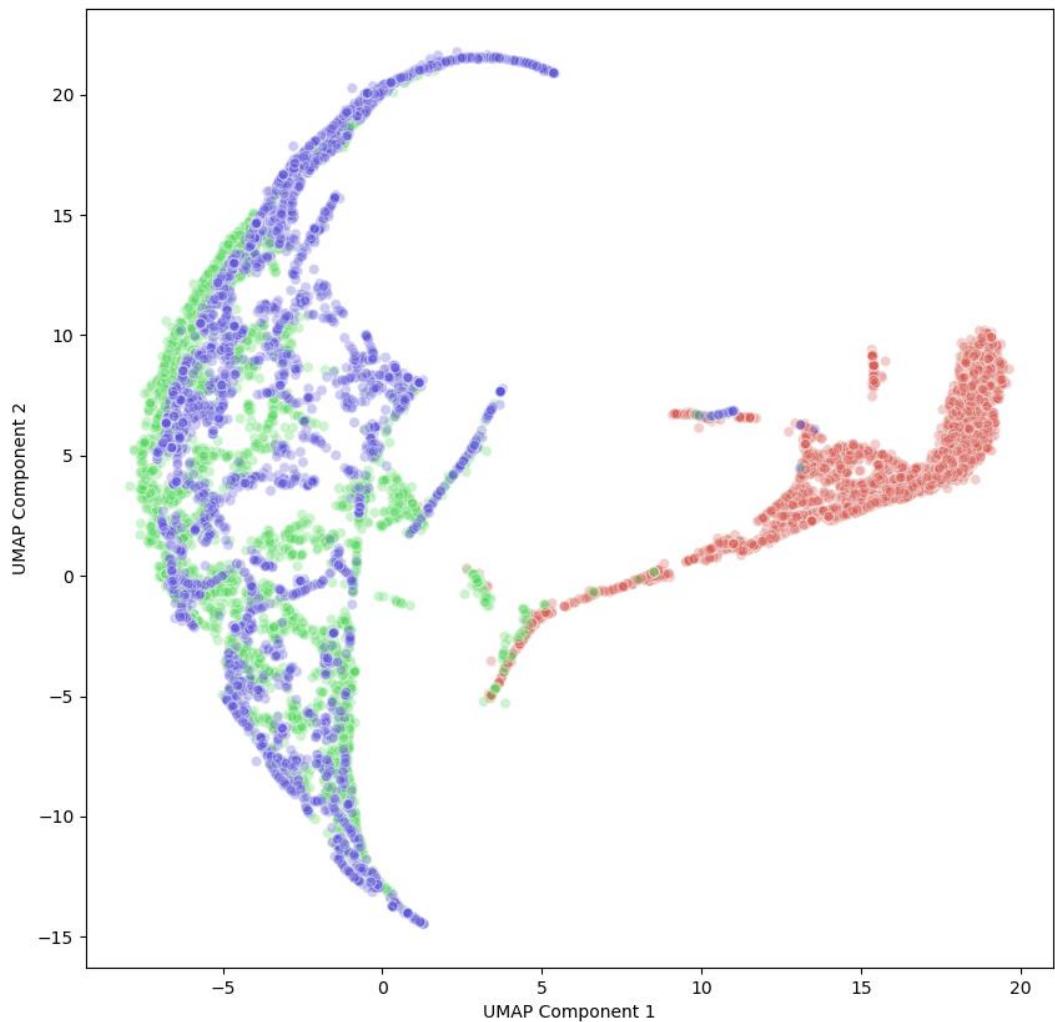


Figure 4.14. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 200. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

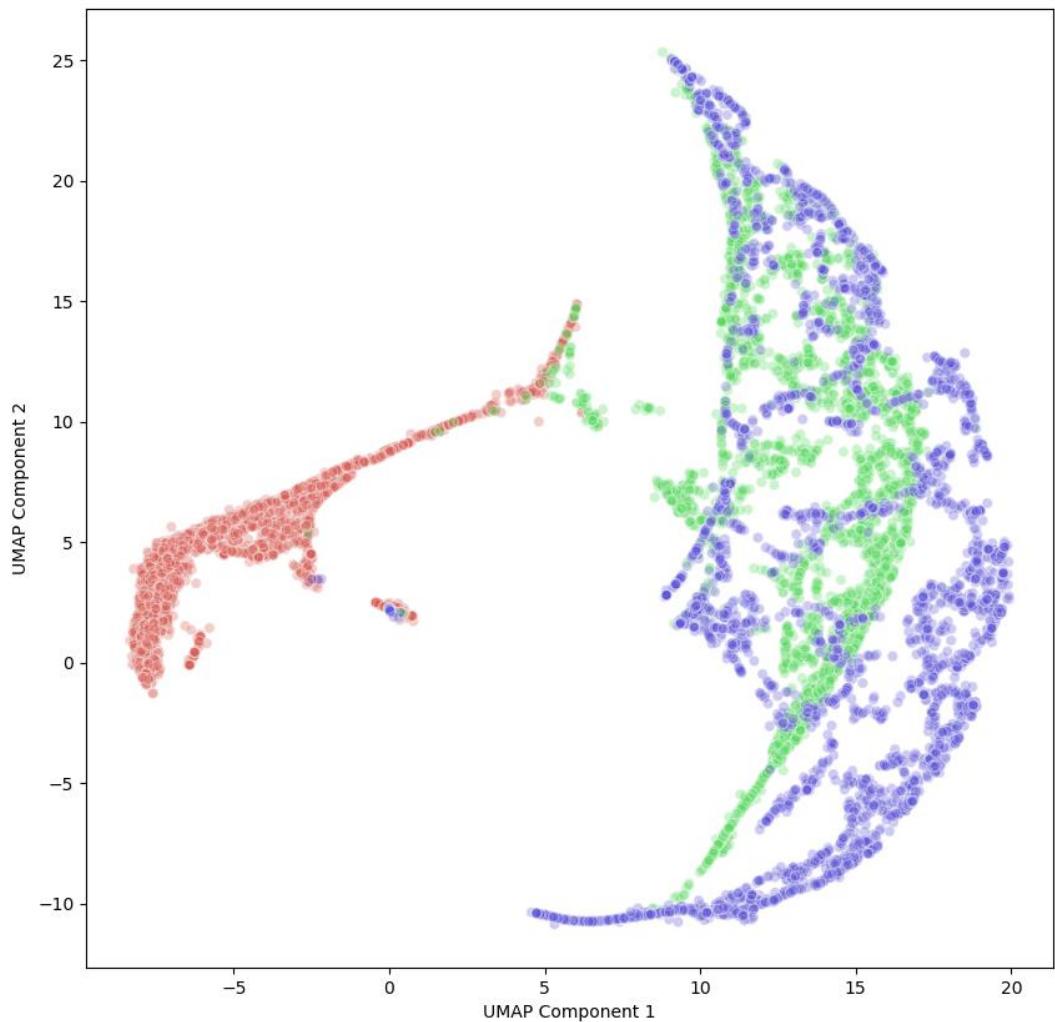


Figure 4.15. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 15; minimum distance = 0.1; transform queue size = 400. Classification labels *small*, *medium*, *large* represented by *red*, *green*, *blue*, respectively.

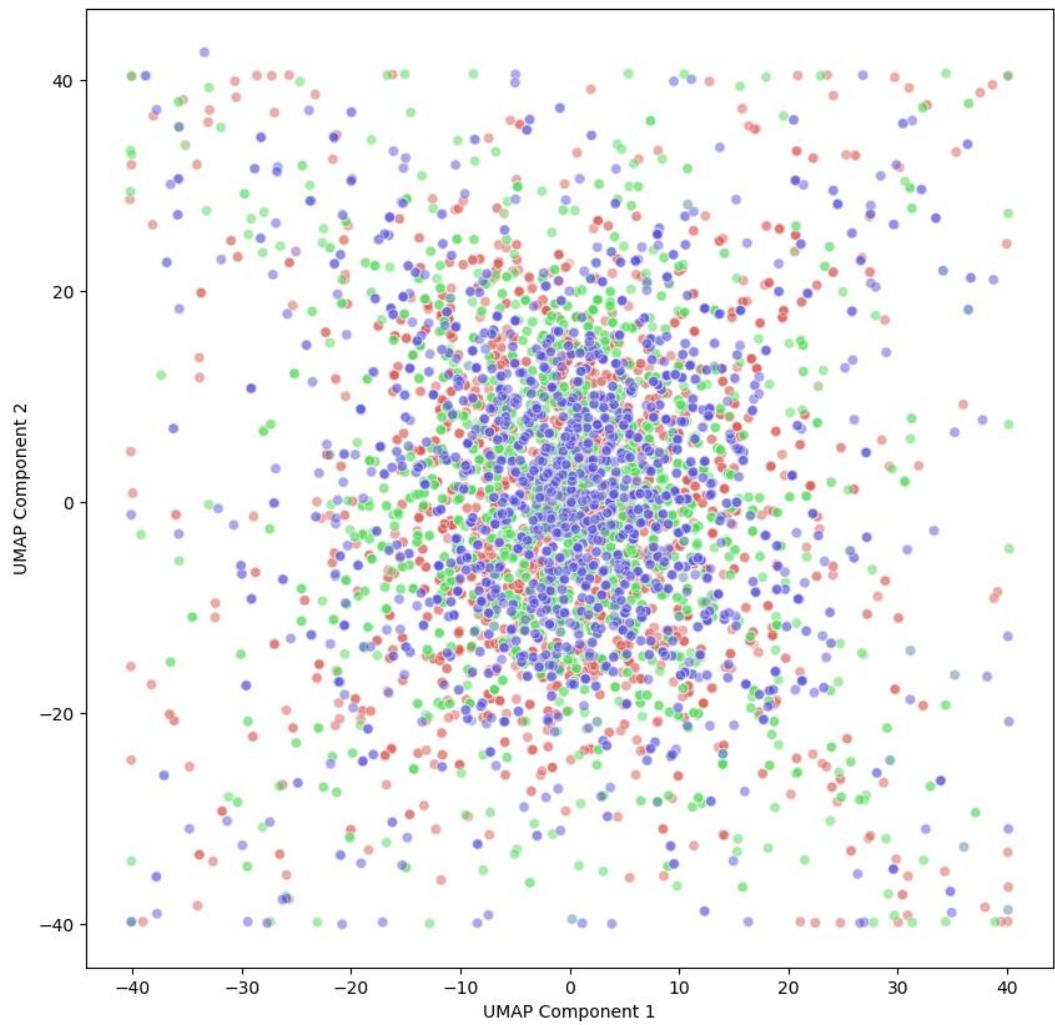


Figure 4.16. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 2; minimum distance = 0.0; transform queue size = 4. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

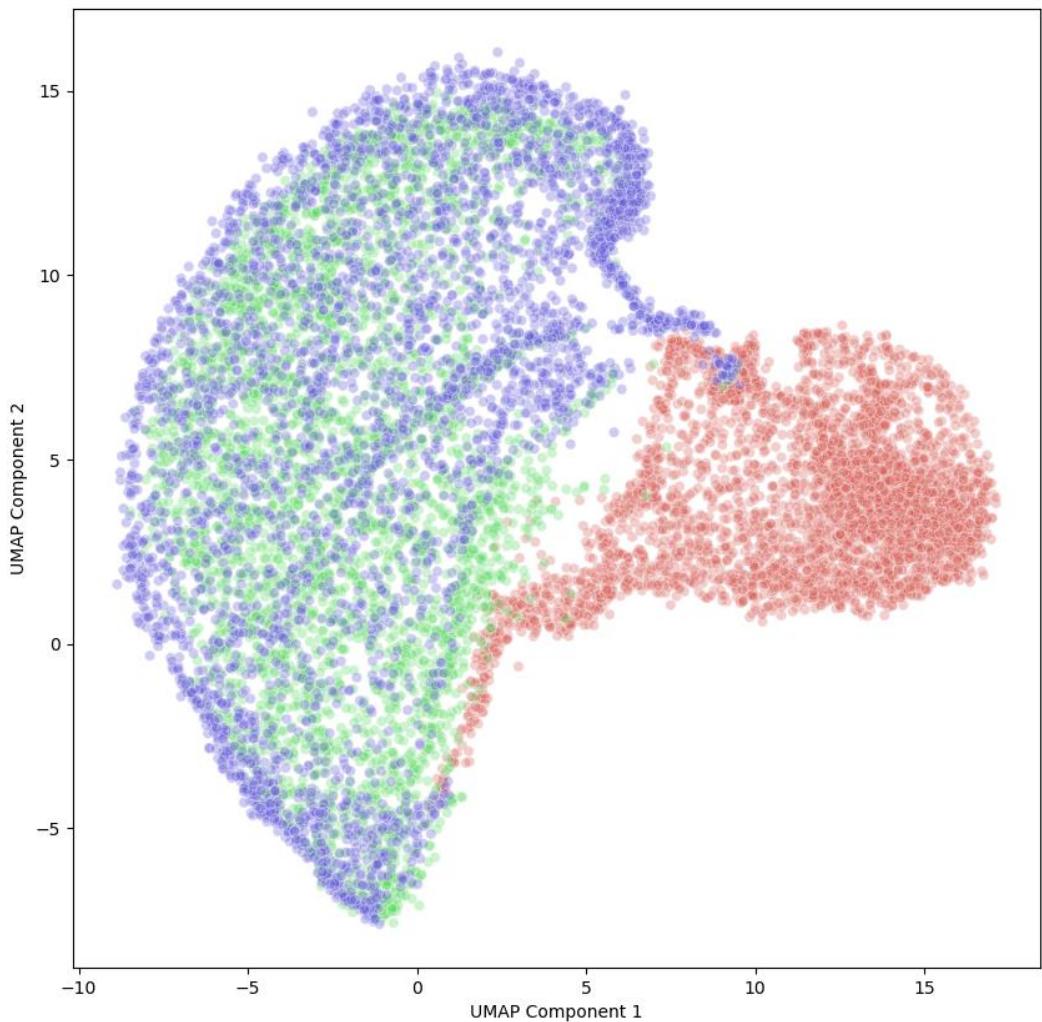


Figure 4.17. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **100**; minimum distance = **0.99**; transform queue size = **400**. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

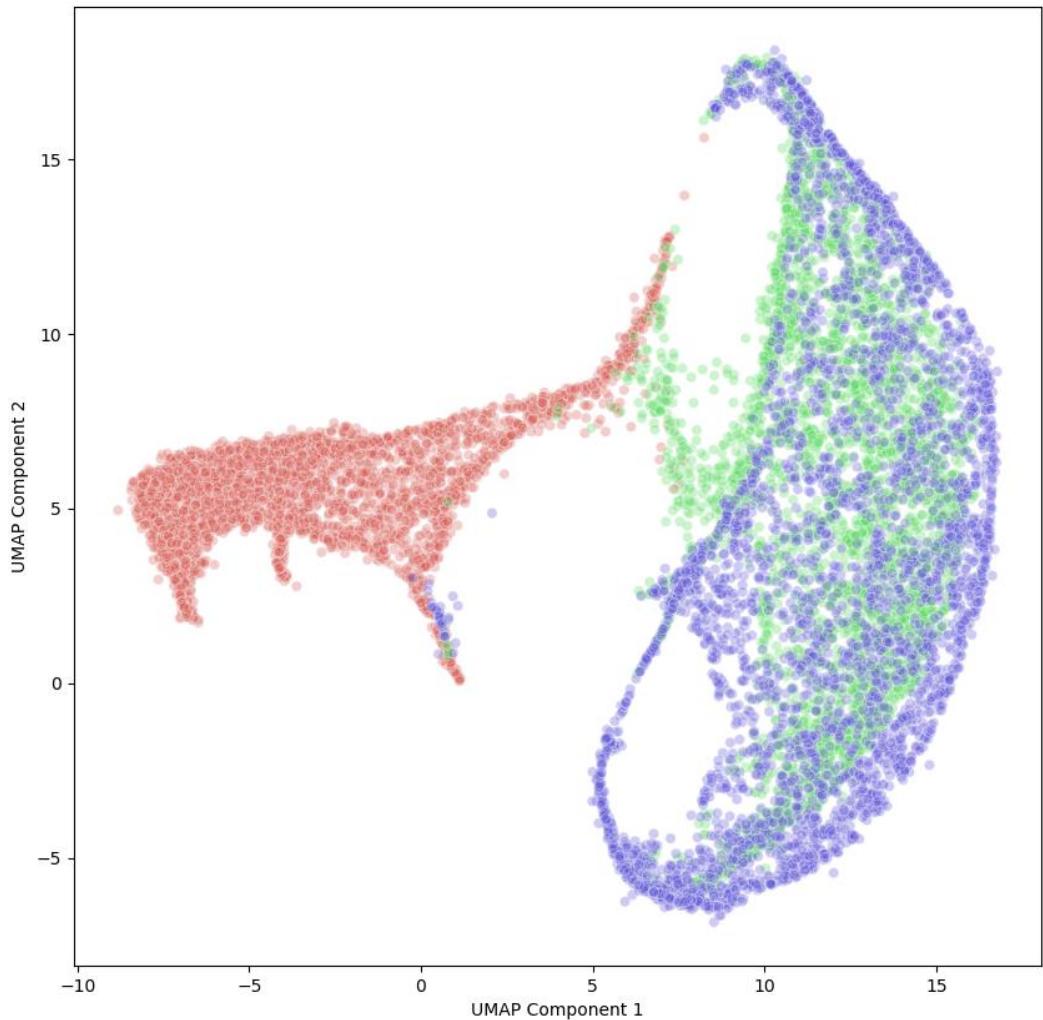


Figure 4.18. UMAP visualisation: components = 3; spread = 1; pairwise distance metric = Euclidean; learning rate = 10; neighbours = 50; minimum distance = 0.50; transform queue size = 200. Classification labels *small*, *medium*, *large* represented by **red**, **green**, **blue**, respectively.

## 5 t-Distributed Stochastic Neighbour Embedding, *t-SNE (Re-classified)*

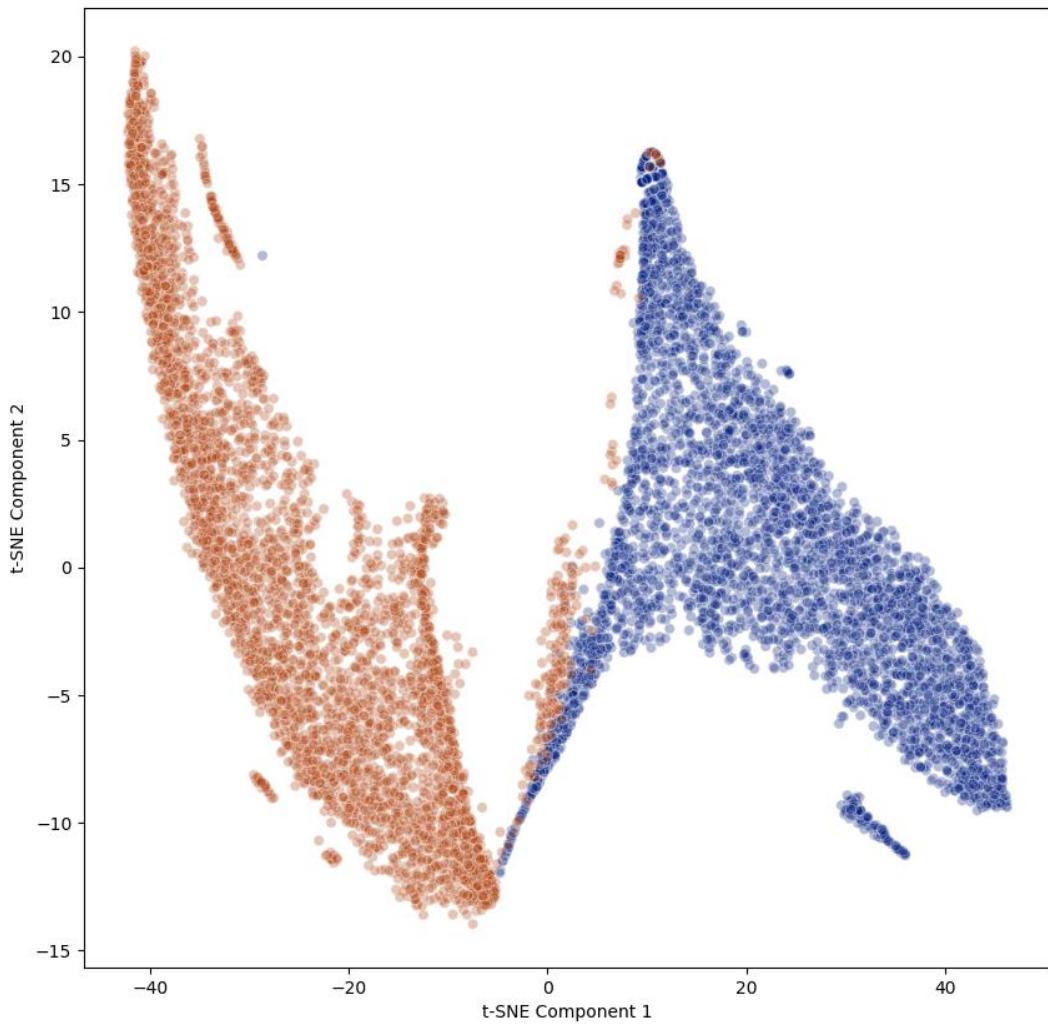


Figure 5.1. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **10**, perplexity = **400**; early exaggeration = **60**; iterations = **5000**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

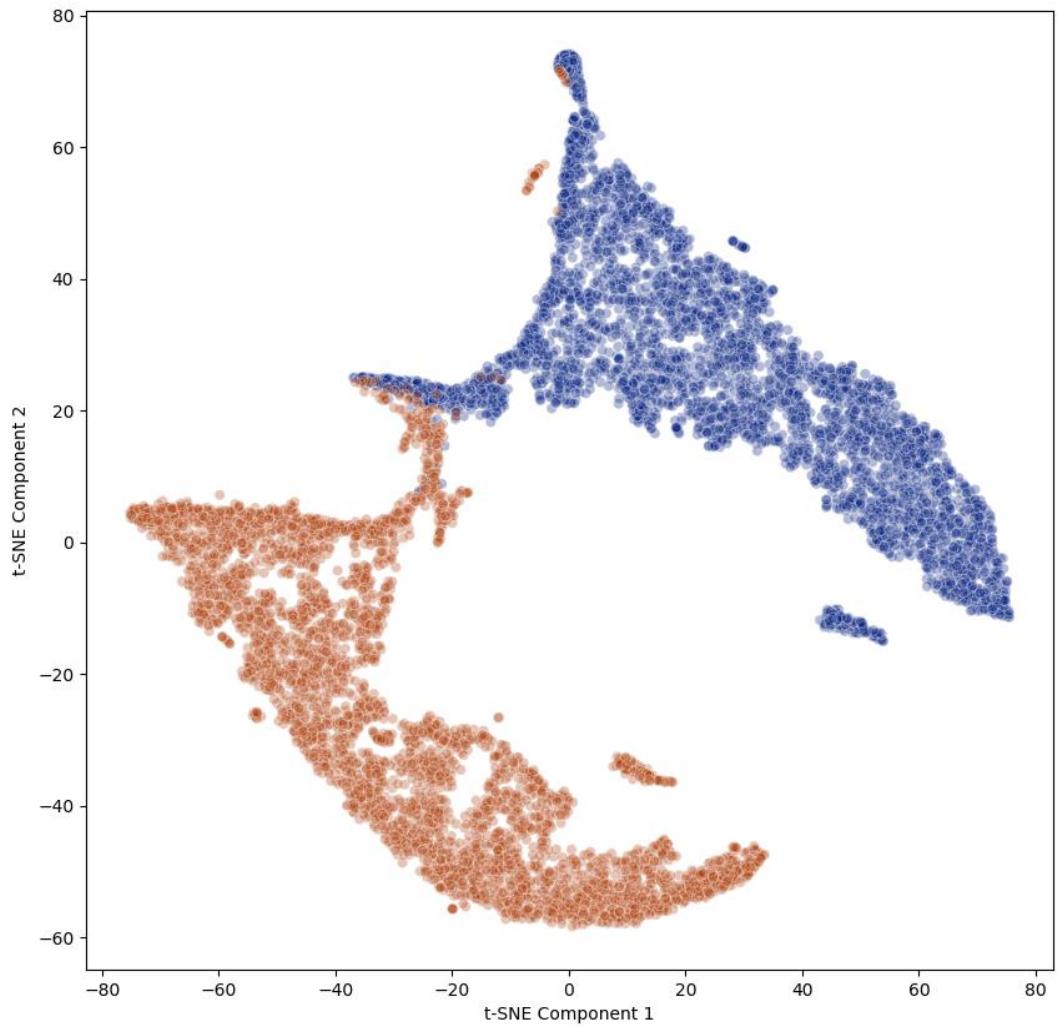


Figure 5.2. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **10**, perplexity = **110**; early exaggeration = **60**; iterations = **5000**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

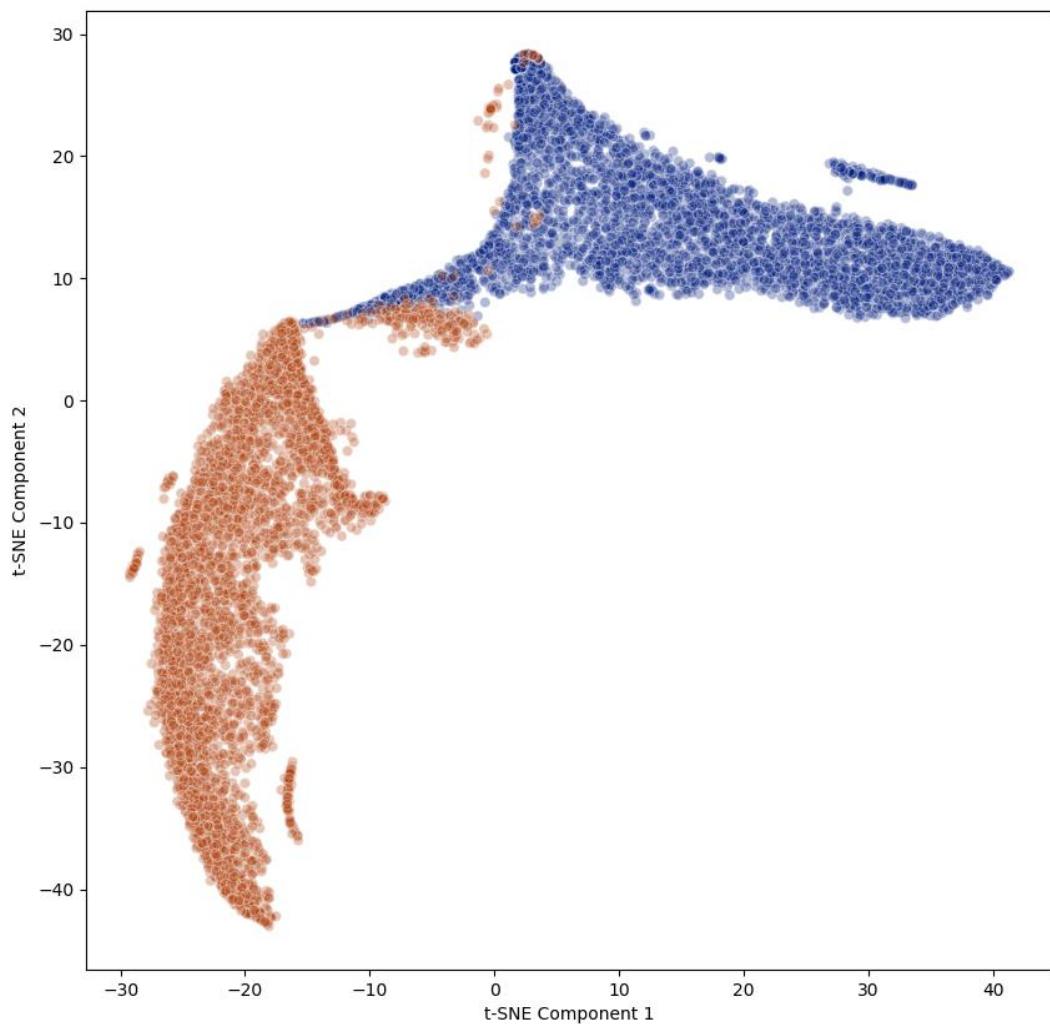


Figure 5.3. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **100**, perplexity = **400**; early exaggeration = **60**; iterations = **5000**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

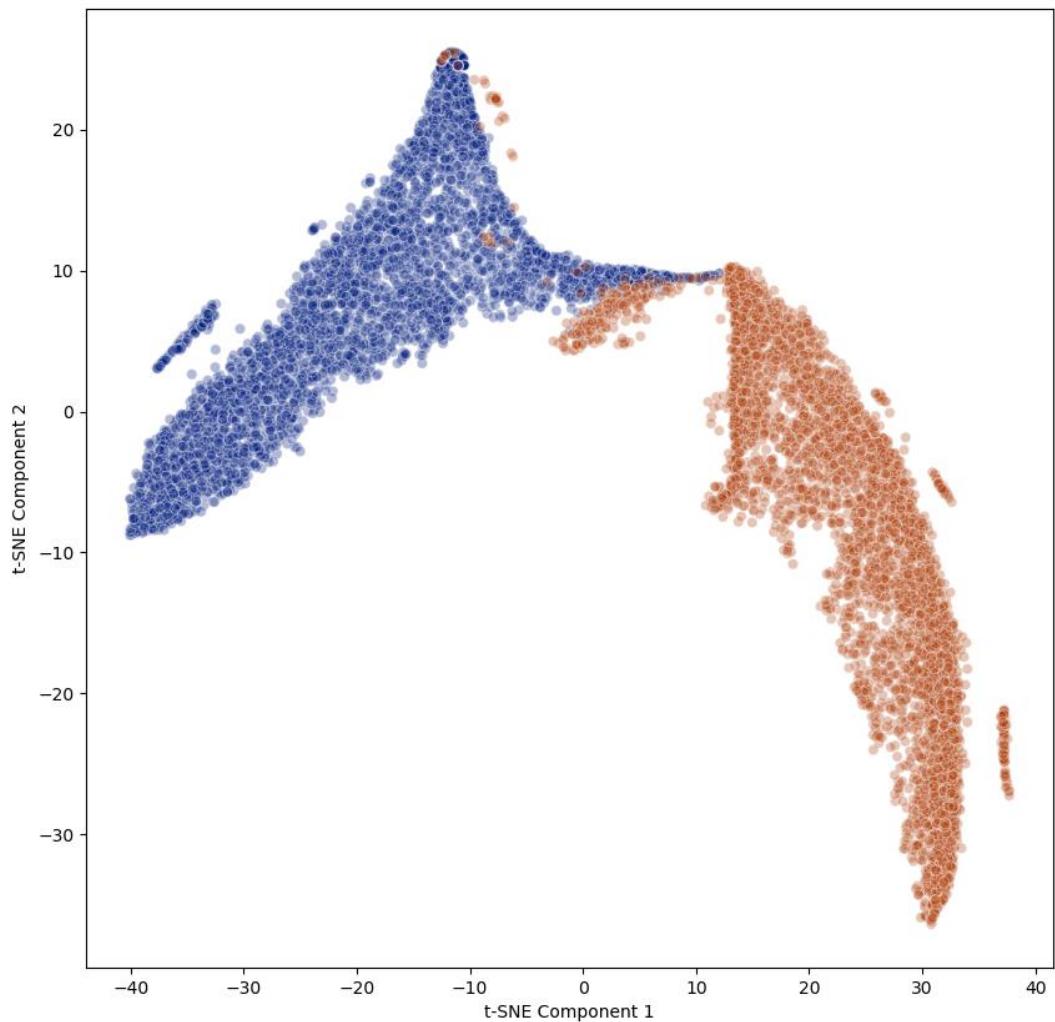


Figure 5.4. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = 50, perplexity = 400; early exaggeration = 60; iterations = 5000. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

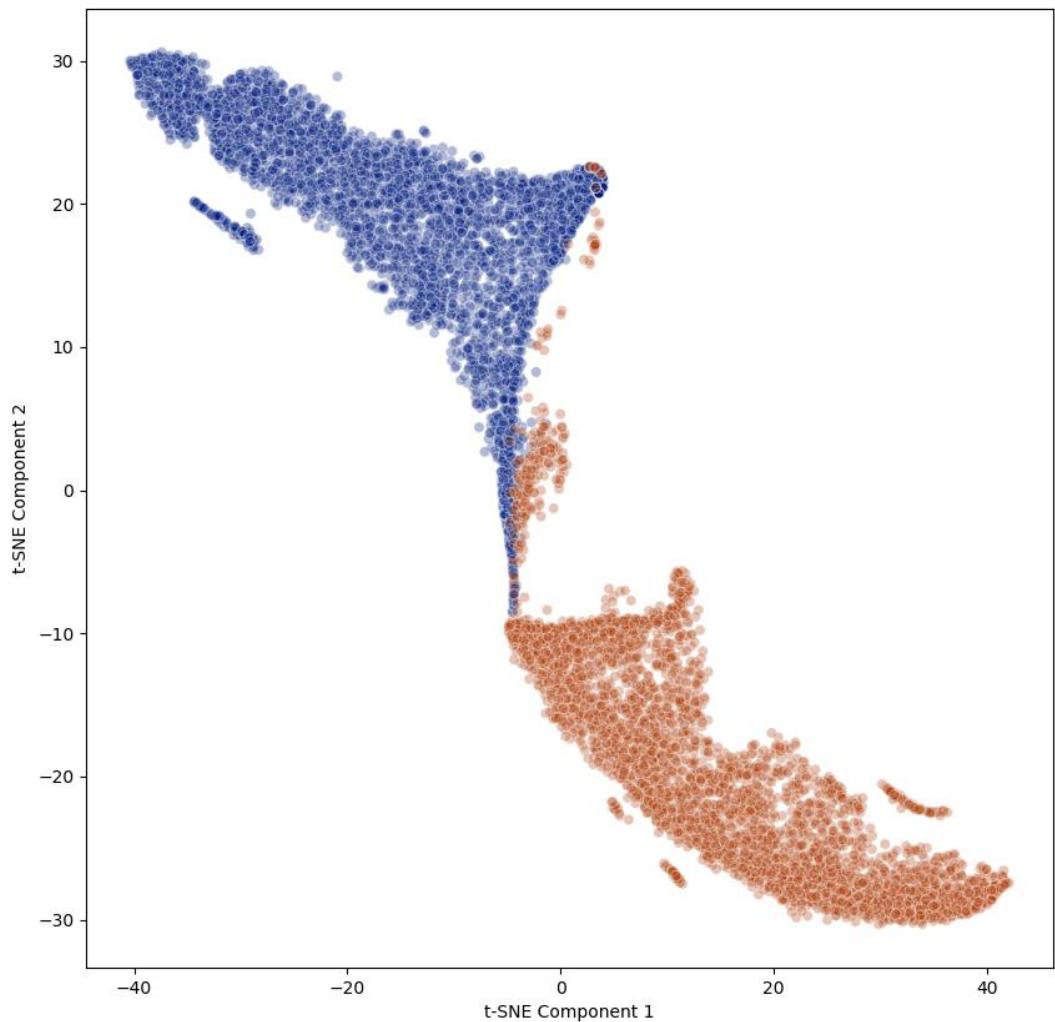


Figure 5.5. t-SNE *re-classified* visualisation: components = 2; gradient calculation method = *Barnes-Hut*; pairwise distance metric = *Euclidean*; learning rate = **50**, perplexity = **400**; early exaggeration = **60**; iterations = **5000**; Barnes-Hut  $\theta = 0.0$ . Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

## 6 Uniform Manifold Approximation and Projection, UMAP (Re-classified)

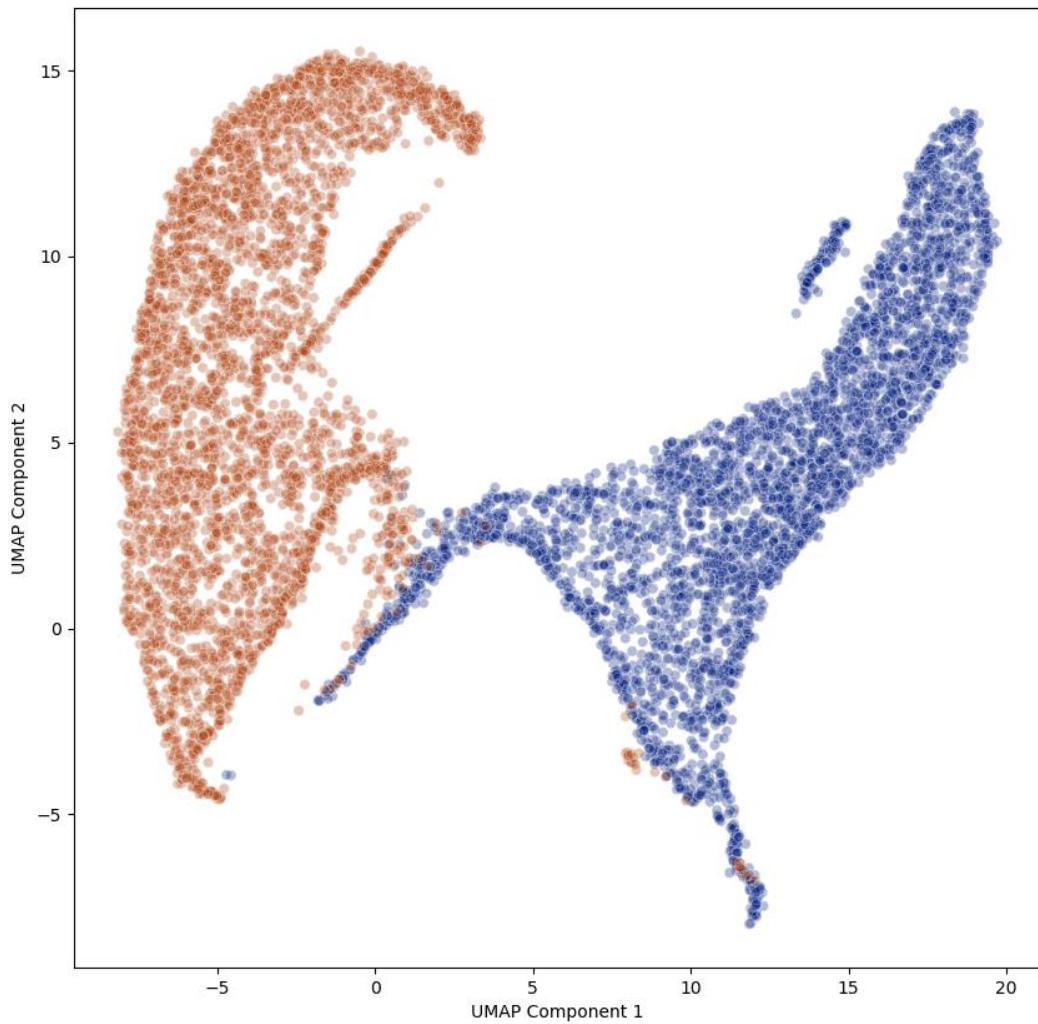


Figure 6.1. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 100; minimum distance = 0.50; transform queue size = 200. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

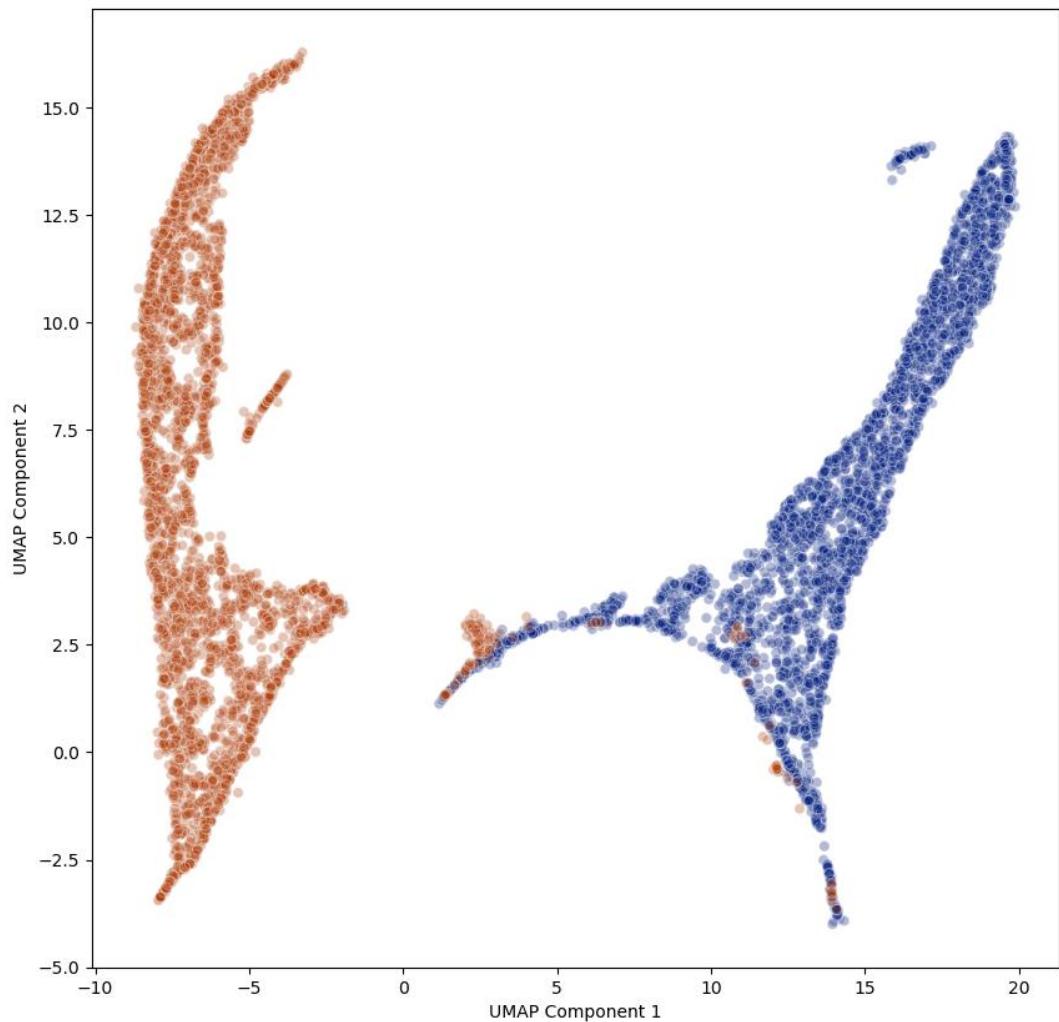


Figure 6.2. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = **100**; minimum distance = **0.10**; transform queue size = **400**. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

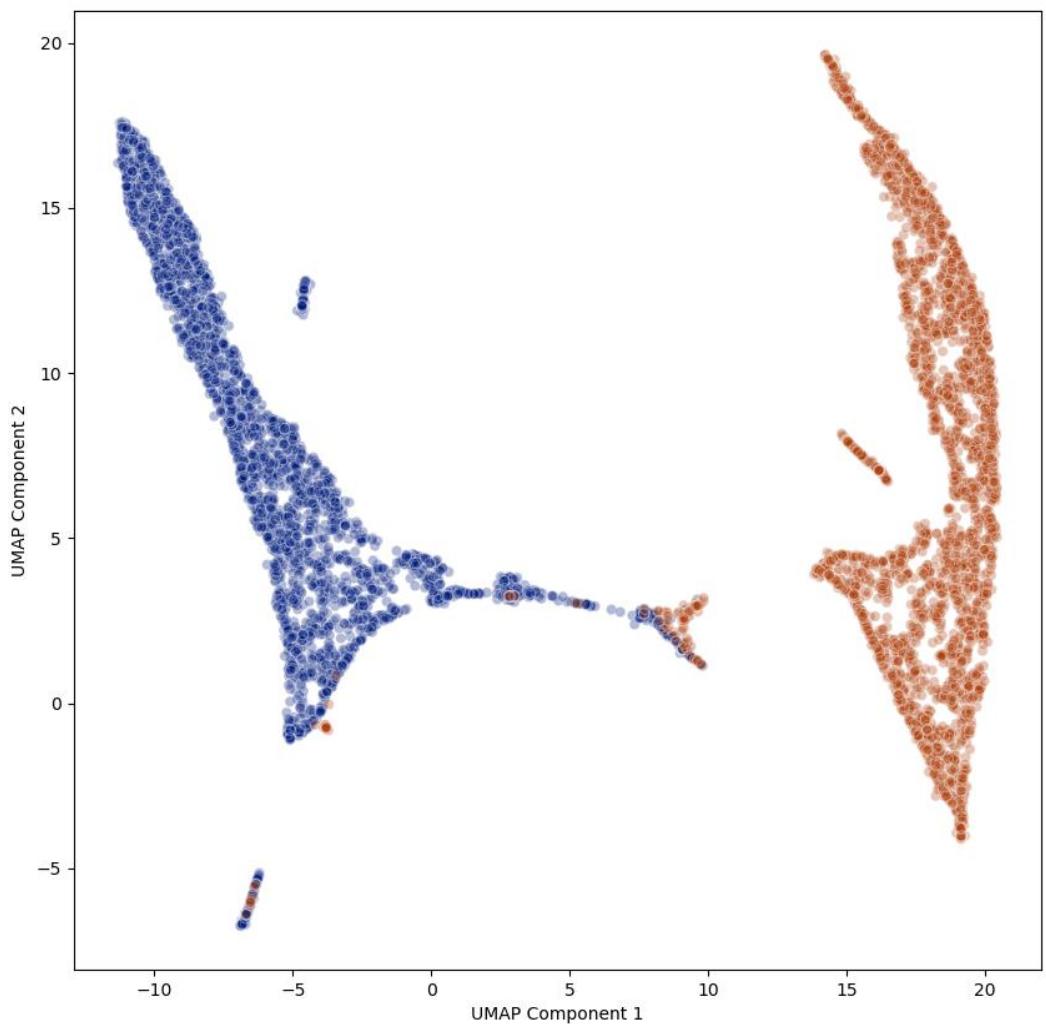


Figure 6.3. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 50; minimum distance = 0.10; transform queue size = 400. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

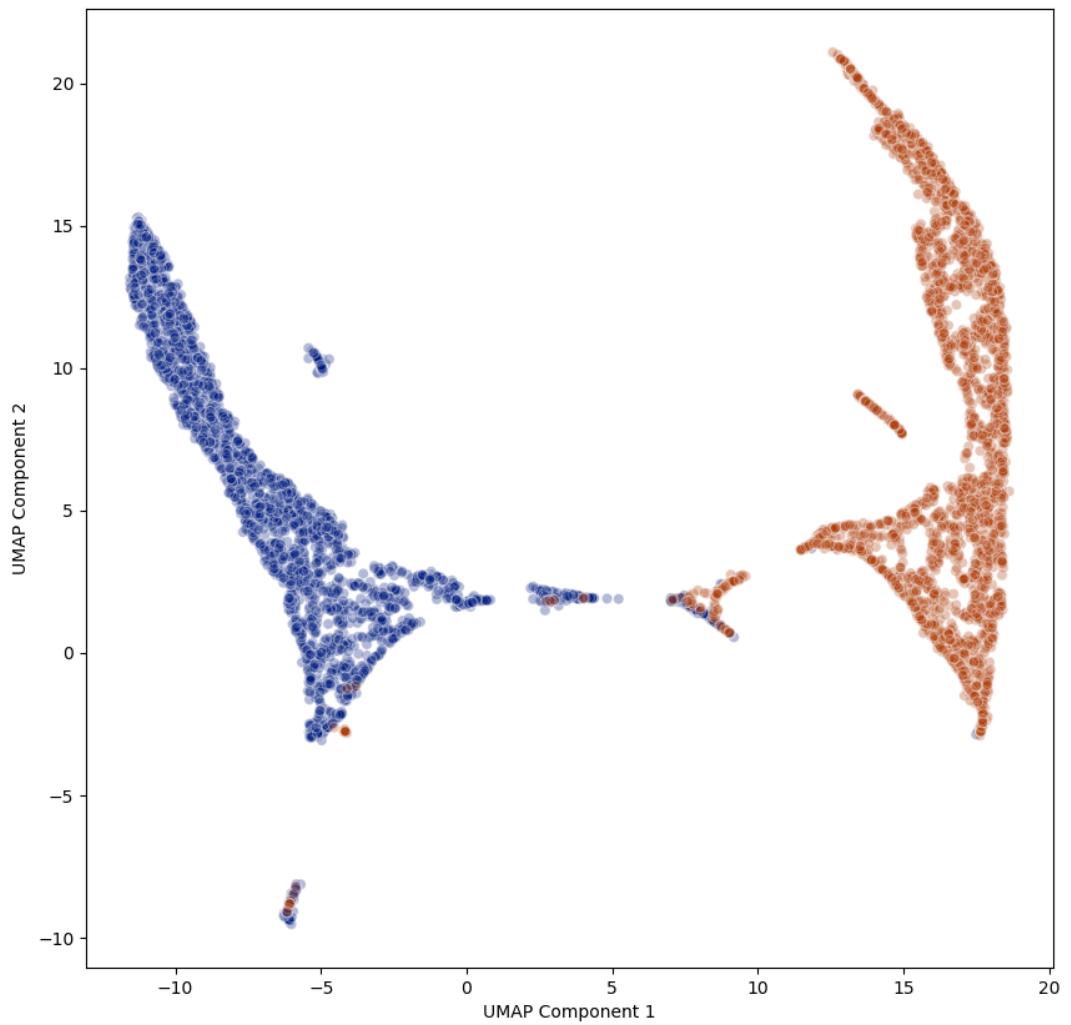


Figure 6.4. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 50; minimum distance = 0.05; transform queue size = 800. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.

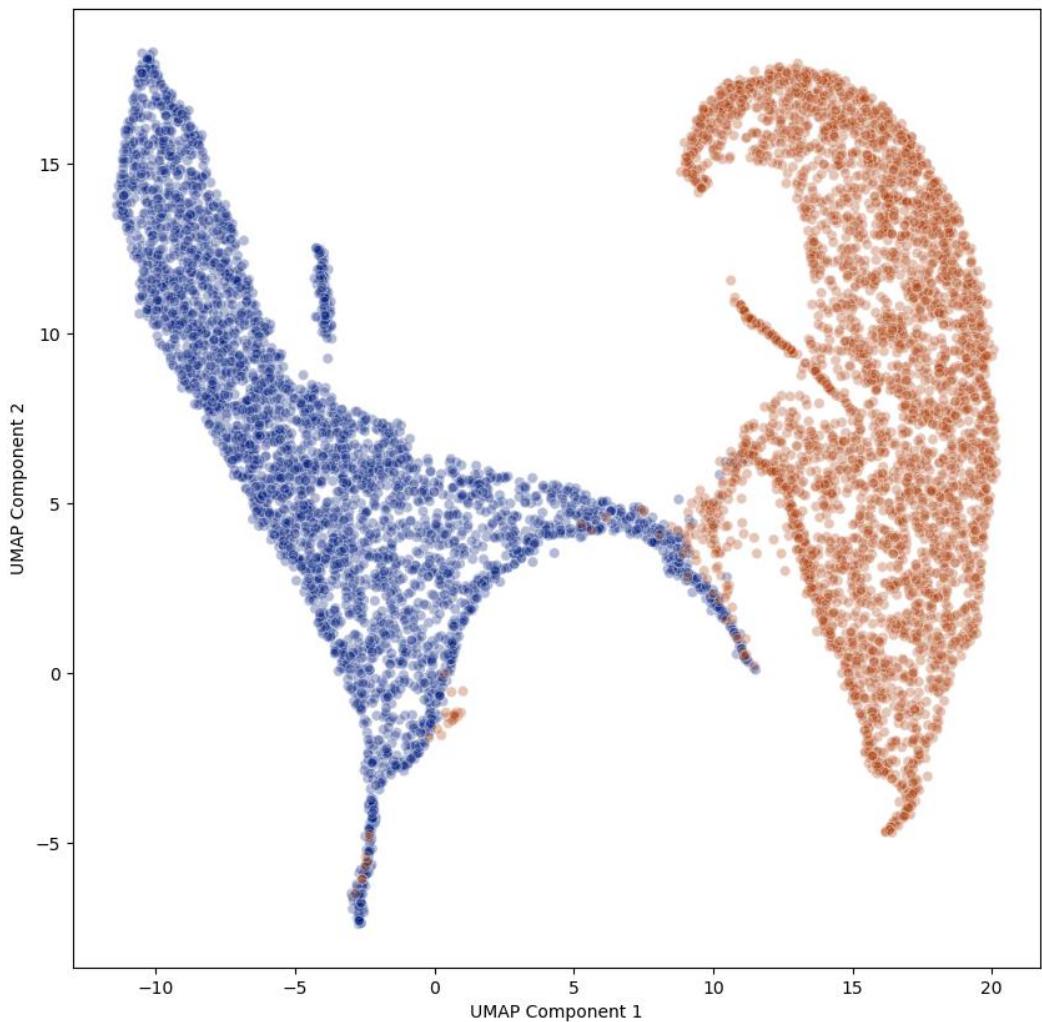


Figure 6.5. UMAP visualisation: components = 2; spread = 1; pairwise distance metric = *Euclidean*; learning rate = 10; neighbours = 50; minimum distance = 0.50; transform queue size = 800. Classification labels *small* and *large* represented by *blue* and *orange*, respectively.