

The Department of Computer Science

**CIS4118 – Coursework 2**

Student: Oliver Pallis – 22917390

Date: 15/05/2023

Contents

[1. Introduction 4](#_Toc135038000)

[2. Background 4](#_Toc135038001)

[3. Analysis 4](#_Toc135038002)

[4. Code 5](#_Toc135038003)

[5. Results 7](#_Toc135038004)

[6. Evaluation 13](#_Toc135038005)

[7. Summary 13](#_Toc135038006)

[8. References 15](#_Toc135038007)

[9. Appendices 17](#_Toc135038008)

[Figure 1 - t-SNE Visualisation of HAPT Dataset with Categorically Coloured Activity Labels 7](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038009)

[Figure 2 - t-SNE Visualisation of HAPT Dataset with Random Forest Predicted Labels 8](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038010)

[Figure 3 - t-SNE 2 Confusion Matrix 1 9](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038011)

[Figure 4 - t-SNE Visualisation of HAPT Dataset with True Activity Coloured Labels 9](#_Toc135038012)

[Figure 5 - PCA Visualisation of HAPT Dataset with Categorically Coloured Activity Labels 10](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038013)

[Figure 6 - PCA Visualisation of HAPT Dataset with Random Forest Predicted Labels 11](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038014)

[Figure 7 - PCA 2 Confusion Matrix 12](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038015)

[Figure 8 - PCA Visualisation of HAPT Dataset with True Activity Coloured Labels 12](file:///C:\Users\Ocpal\OneDrive\Documents\Edge%20Hill\Data%20Visualisation\CW2\FINAL\CW2.docx#_Toc135038016)

[Figure 9 - t-SNE Visualisation of the Walking Activity with True Labels 17](#_Toc135038017)

[Figure 10 - t-SNE Visualisation of the Walking\_Upstairs Activity with True Labels 18](#_Toc135038018)

[Figure 11 - t-SNE Visualisation of the Walking\_Downstairs Activity with True Labels 19](#_Toc135038019)

[Figure 12 - t-SNE Visualisation of the Sitting Activity with True Labels 20](#_Toc135038020)

[Figure 13 - t-SNE Visualisation of the Standing Activity with True Labels 21](#_Toc135038021)

[Figure 14 - t-SNE Visualisation of the Laying Activity with True Labels 22](#_Toc135038022)

[Figure 15 - t-SNE Visualisation of the Stand\_to\_Sit Activity with True Labels 23](#_Toc135038023)

[Figure 16 - t-SNE Visualisation of the Sit\_to\_Stand Activity with True Labels 24](#_Toc135038024)

[Figure 17 - t-SNE Visualisation of the Sit\_to\_Lie Activity with True Labels 25](#_Toc135038025)

[Figure 18 - t-SNE Visualisation of the Lie\_to\_Sit Activity with True Labels 26](#_Toc135038026)

[Figure 19 - t-SNE Visualisation of the Stand\_to\_Lie Activity with True Labels 27](#_Toc135038027)

[Figure 20 - t-SNE Visualisation of the Lie\_to\_Stand Activity with True Labels 28](#_Toc135038028)

[Figure 21 - PCA Visualisation of the Walking Activity with True Labels 29](#_Toc135038029)

[Figure 22 - PCA Visualisation of the Walking\_Upstairs Activity with True Labels 30](#_Toc135038030)

[Figure 23 - PCA Visualisation of the Walking\_Downstairs Activity with True Labels 31](#_Toc135038031)

[Figure 24 - PCA Visualisation of the Sitting Activity with True Labels 32](#_Toc135038032)

[Figure 25 - PCA Visualisation of the Standing Activity with True Labels 33](#_Toc135038033)

[Figure 26 - PCA Visualisation of the Laying Activity with True Labels 34](#_Toc135038034)

[Figure 27 - PCA Visualisation of the Stand\_to\_Sit Activity with True Labels 35](#_Toc135038035)

[Figure 28 - PCA Visualisation of the Sit\_to\_Stand Activity with True Labels 36](#_Toc135038036)

[Figure 29 - PCA Visualisation of the Sit\_to\_Lie Activity with True Labels 37](#_Toc135038037)

[Figure 30 - PCA Visualisation of the Lie\_to\_Sit Activity with True Labels 38](#_Toc135038038)

[Figure 31 - PCA Visualisation of the Stand\_to\_Lie Activity with True Labels 39](#_Toc135038039)

[Figure 32 - PCA Visualisation of the Lie\_to\_Stand Activity with True Labels 40](#_Toc135038040)

# Introduction

The following document is a report that is based on the Human Activities and Postural Transitions (HAPT) dataset and how visualisations can be used to extract meaning from the dataset. This will be accomplished by creating plots to visualise the true data itself, and plots showing how machine learning algorithms can be applied to the data set. Background information on this specific subject will be investigated by reviewing similar works and which methods were used to achieve best results. With the methods identified, the code was then created to implement identified methods to manage the large dataset, understand the data, train algorithms on the data, produce visualisations that capture the meaning behind the HAPT dataset and also identify relationships between data points. This report will highlight the steps that were taken to accomplish this task, discuss the reasoning why certain methods were chosen over others, and what can be interpreted from the visualisations.

# Background

The HAPT dataset consists of motion data that was recorded using smartphones on 30 volunteers, aging between 19 and 48 years old. The 30 subjects were required to perform 12 basic activities with 3 of them being dynamic, 3 of them being static and 6 of them being transitional. The 3 dynamic activities were walking, walking upstairs, and walking downstairs. The 3 static activities were standing, sitting, and lying. The 6 transitional activities were stand to sit, sit to stand, sit to lie, lie to sit, stand to lie and lie to stand. The data was collected by using the accelerometer and gyroscope sensors that are built into the smartphones. A total of 10,229 observations were recorded from this experiment and each observation contains 561 measurements of different features, creating a large and highly complexed dataset. The HAPT dataset falls under Human Activity Recognition (HAR), which consists of the recognising, interpreting, and assessing human activities, such as static, dynamic and transitional movements (Irfan et al., 2021). Furthermore, machine learning algorithms can use data sets from HAR for classification purposes to provide insights into the data, allowing for better understandings and even foresight. The machine learning algorithms can be used to automatically identify patterns and categorise activities based on the sensor data. Therefore, particularly for machine learning specialists and for researchers in HAR, the HAPT dataset is an excellent resource for testing machine learning models, due to the vast amount of observations and characteristics of the data (Rahimi Taghanaki, Rainbow and Etemad, 2021). The use of machine learning algorithms, such as Random Forest (RF), Support Vector Machines (SVM) or Decision Trees (DT) can be used and applied to the HAPT dataset to obtain a higher accuracy score in detecting and recognising activities. This is supported by Ramanujam et al. (2021), who writes that in addition to classification purposes, feature selection or dimensionality reduction and visualisations can also be used to improve the effectiveness of the model. Feature selection can be used on the data set to reduce the amount of features and increases the accuracy of the classification model. According to Irfan et al. (2021), this is also known as dimensionality reduction, which are techniques such as t-Distributed Stochastic Neighbour Embedding (t-SNE) or Principal Component Analysis (PCA), the machine learning model’s performance can be increased by reducing the number of features. This can allow for patterns in the data or relationships between classes, or activities, to be identified and then presented in a visualisation. Therefore, the results that can be obtained from the HAPT dataset could be used to further develop the capabilities and possibilities of machine learning algorithms in combination with HAR data or research.

# Analysis

Given the nature and purpose of the HAPT dataset, machine learning algorithms can be utilised to carry out classification tasks to be able to determine which measurements belong in which activity category. Furthermore, through the use of visualisations, insights can be obtained regarding the similarities between movements and how relationships between the movements can be identified. This can be accomplished with the use of scatter plots, where measurements can be visualised as clusters, which could indicate how the different subjects, or participants, might complete certain movements. In addition, the visualisations can also suggest which measurements may contain similarities between the different movement activities. According to Maaten and Hinton (2008), t-Distributed Stochastic Neighbour Embedding (t-SNE) is a dimensionality reduction algorithm and is useful for capturing and visualising complex high dimensional data sets. This is often used alongside machine learning algorithms such as RF, DT or KNN. Irfan et al. (2021), who has worked with the HAPT dataset with machine learning had found that the Random Forest algorithm had produced good results for classification tasks. This was also proved in research carried out by Ramanujam et al. (2021), who was able to receive an accuracy score of 96% when using the RF classifier on the HAPT data set with 561 features. The RF classifier had a higher performance value than other classifiers, such as K-Nearest Neighbour (KNN) and SVM. Therefore, in order to carry out this task, t-SNE and the RF classifier would be used to both create visualisations of the data that would show the measurements of the different activity groups and how they relate to each other, whilst also training a RF algorithm to accurately predict the classes for future data sets. In addition to this, clustering feature extraction along with classification would also be used by combining PCA with RF to compare results from the two combinations. This combination was experimented with in a study by Ballı, Sağbaş and Peker (2019), who had also found that the RF classifier to be successful in classifying human activities.

# Code

A total of six source code files were created to address this task, where three were created using t-SNE and three for PCA. Each source code in this section will be introduced and discussed in the order of which the plots will be shown in Results of this report, as this was the order of how the codes were tested and evaluated. Furthermore, each source code was created following the PEP8 standards. It is important to note that the HAPT data set has been split into training and testing data, with 70% being the training data, and 30% being the testing data. The original data has not been altered in any way, and the code has been designed to work with the data in its original format. Although, given the split of the data it was found that there were measurements for 30 participants in the training data, but only measurements for 9 of the participants were present in the testing data.

The first source code, titled “t-SNE 1”, that was created produces a visualisation of the HAPT dataset by using t-SNE to visualise the high-dimensional data in a lower-dimensional space and in this case, a 2D space. The code firstly loads the training data, the test data, the subject IDs for both train and test, and also the activity labels. By using StandardScaler, the data is scaled and then filtered. Afterwards, t-SNE is applied to the scaled training and test data, creating two-dimensional embeddings of the data points within the data set. Afterwards, a scatter plot is created of the embeddings with the three colours assigned to the data points, depending on which activity label the data points belong to. Whilst this code is not performing classification or feature extraction, the plot was created for informational purposes as the plot provides a visual representation of the data set and which data points belong to which activity category. Furthermore, the plot will show how the clusters of the activity categories are present within the dataset.

The second source code, titled “t-SNE”, produces a plot of the HAPT dataset and uses the RF classifier to predict the activity labels for the data points. The code follows the same first steps as t-SNE 1by loading the HAPT data files and is then scaled. T-SNE is then applied to the scaled training and test data to reduce the dimensionality to two dimensions, where the t-SNE embeddings for the testing data is then extracted. Afterwards, a parameter grid is created for the RF classifier, which is then built. To find the best hyperparameters for the RF classifier, a grid search is performed and once identified, the RF classifier uses the best hyperparameters to be trained on the data set. The accuracy of the classifier and the confusion matrix is then calculated and printed, to evaluate the performance of the classifier. This produces the scatter plot, visualising the t-SNE embeddings with the labels predicted by the RF classifier. The plot shows twelve different colours for the twelve different activities respectively.

The third source code, titled “t-SNE 3”, is a near replica of the first source code, with the only main difference being that two lists were created for the static and dynamic activities to be used to assign colours to the different activities. A list of colours is also created with a pre set range of twelve colours assigned. The purpose of this code is to produce a visualisation of the HAPT data set but with the true labels for each activity, rather than the predicted labels or the categorical labels. This plot can also be used to visually compare the plots from the predicted labels and the true labels, whilst also providing an accurate visualisation of the clusters and relationships between each activity. Furthermore, additional plots are also created for each activity alone, highlighting potential outliers or anomalies in the data.

The fourth source code, titled “PCA 1”, works similarly to the first source code, but with PCA applied instead of t-SNE. The HAPT data files are loaded and then scaled, with the dynamic and statis activities also defined. This was necessary to assign the colours to the activities for the scatter plot that would be produced. After the data is scaled, PCA is applied to the combined scaled data to reduce the dimensionality to two dimensions. The transformed test data is then separated from the combined data and fitted into a data frame with the true labels and the subject IDs. Afterwards, the scatter plot is created with the assigned colours used for the three movement categories, dynamic, static and transitional. This creates a visualisation of the HAPT data set, showing how the different categories hold the data points in the two-dimensional space. As used in the first plot, the colours red, green, and blue were chosen as these colours can still be seen in the visualisation should overlapping occur.

The fifth source code, titled “PCA 2”, follows the same outline as the second source code, but with PCA combined RF instead. The code applies PCA to reduce the dimensionality of the data to two dimensions, performs the grid search using cross-validation to determine which are the best hyperparameters would be best used for the RF classifier. Using those best hyperparameters, the RF classifier is then trained on the dataset, and then used to predict the activity labels on the HAPT dataset. The confusion matrix is calculated and printed along with the accuracy score to evaluate how well the classifier had performed. The final result is a scatter plot of the PCA embeddings along with the RF predictions of activity labels.

The sixth and final source code, titled “PCA 3”, creates a visualisation of the HAPT dataset with the PCA embeddings, but with the true labels rather than the predicted labels, and with the true labels for each activity rather than the three activity categories. With this, and the previous PCA visualisations, the plots created allow for the HAPT data set to be visualised, with the PCA embeddings, showing the true activity labels for the three categories, individual activities with true labels, and also the individual activities with the RF predictions. There are also additional plots with the true labels for each activity to allow for further visual insight as to how each activity is measured.

Each source code was created with the purpose of allowing the visualisations to be interpreted and the different methods to be compared with each other. As the HAPT data set can be utilised in many different ways, these codes show a few ways that the data can be utilised and the different methods that were used could be assessed.

# Results

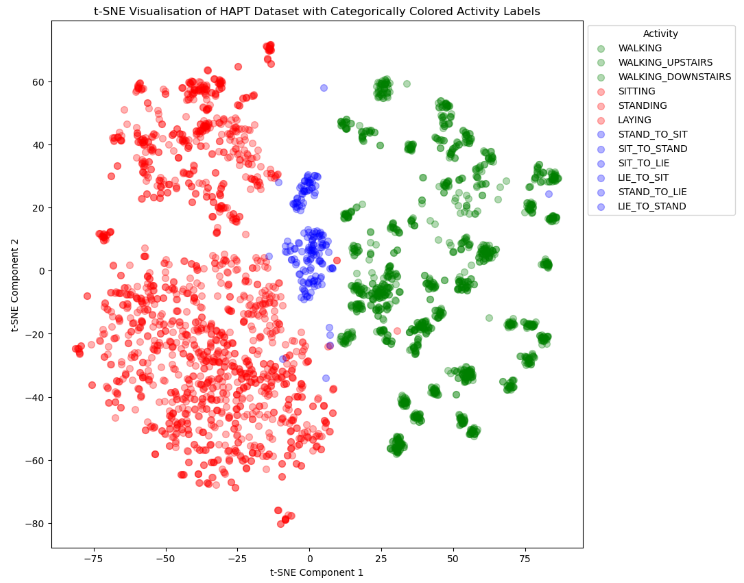
The following plot presents the t-SNE embeddings of the HAPT dataset with the three colours representing the three categories of activities that were carried out by the participants, namely static, dynamic, and transitional movements.

Figure - t-SNE Visualisation of HAPT Dataset with Categorically Coloured Activity Labels

The plot above clearly shows a separation between the three activities with the static movements clearly separated from the dynamic and have numerous clusters within, which could represent measurements made from various age groups. The transitional measurements are found between the dynamic and clusters, as expected, highlighting the transition from a dynamic activity to a static. There is some overlap between the three activity groups, which, to a degree, can be expected with the similarity between some activities. Furthermore, there are some outliers present such as the measurement at 85 along the X and 30 along the Y in figure 1with a transitional measurement found amongst the dynamic measurements, which can represent either an error in the data, a labelling error, or a sudden difference in acceleration. Overall, the plot shows that there is a clear difference in sensor data between the activities and the t-SNE algorithm had successfully learned to distinguish between the different activity types, which could be helpful for further analysis with machine learning models.

The following plot is also shows the results of the dimensionality reduction using t-SNE followed by the Random Forest algorithm. Much like the previous plot, the following plot shows the data points at the same positions but rather than the data points representing the three activity categories, they represent the specific activity to which the movement belongs to.

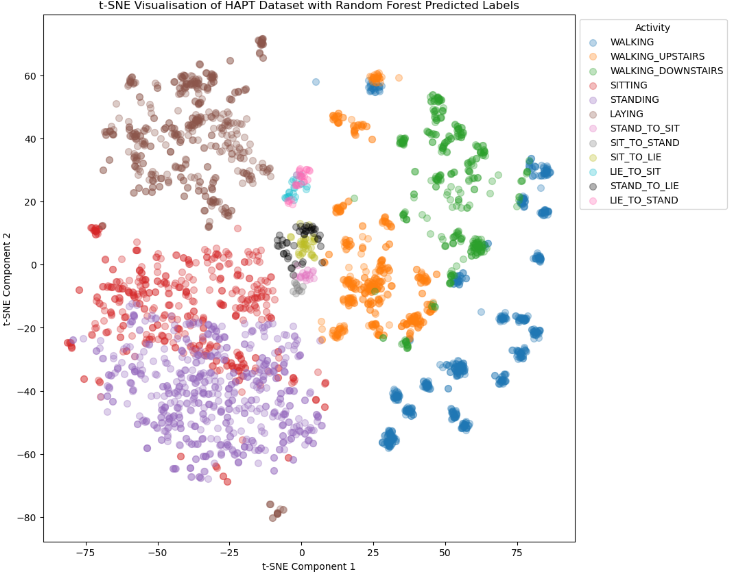


Figure - t-SNE Visualisation of HAPT Dataset with Random Forest Predicted Labels

However, the plot does not show the true labels of the activity but rather the predicted labels determined by the RF classifier. The classifier achieved an accuracy score of 87% and was able to successfully predict and determine the separate classes of data points. The classifier was able to efficiently identify different activities that are clustered closely together, whilst also highlighting some overlapping classes. This indicates that the overlapping activities are harder to distinguish based on the feature set that was used. In addition, this could also be a result of the data complexity or a result of the limitations of t-SNE. However, it is shown that the static movements have measurements that are somewhat similar to each other and are clearly separated from the dynamic movements. Regarding weaknesses of this plot, consider the outlier that was mentioned in figure 1, in which the plot had shown this to be a transitional movement, the RF classifier had predicted this movement to be from the “walking activity”. This is a result of the 13% inaccuracy of the RF classifier, however, the majority of the predictions were true, and can be seen in the following plot. Furthermore, there is a significant overlap in a cluster between the “walking” activity and the “walking\_upstairs” activity, which could be considered as a visible inaccuracy, given that the distance between the cluster and walking is much further than the walking\_upstairs classes. Overall, this t-SNE plot with the RF classifier provided the visualisation that shown the effectiveness of the feature set that was used for the classification, and effectively highlights the different activities, with some activities being harder to differentiate. Below is the confusion matrix for the RF classifier that shows the true positive (TP), false positive (FP), true negative (TN) and the false negative (FN) predictions made by the classifier.

The accuracy is calculated by the RF classifier predicting the y\_labels, or activities, for each measurement and the true predictions are shown on the diagonal elements of the matrix, showing thata 87% of the predictions out of all predictions were correct. In terms of evaluating the model’s performance, the confusion matrix provides information on identifying if something is not right with the model, and can be helpful in locating the source of the issue. In this case, 87% accuracy in A screenshot of a test

Description automatically generated with medium confidenceconsidered efficient and, the 13% inaccuracy can also be seen in the following chart that shows the plot of the data set again, but with the true labels rather than the predicted labels.

Figure - t-SNE 2 Confusion Matrix 1

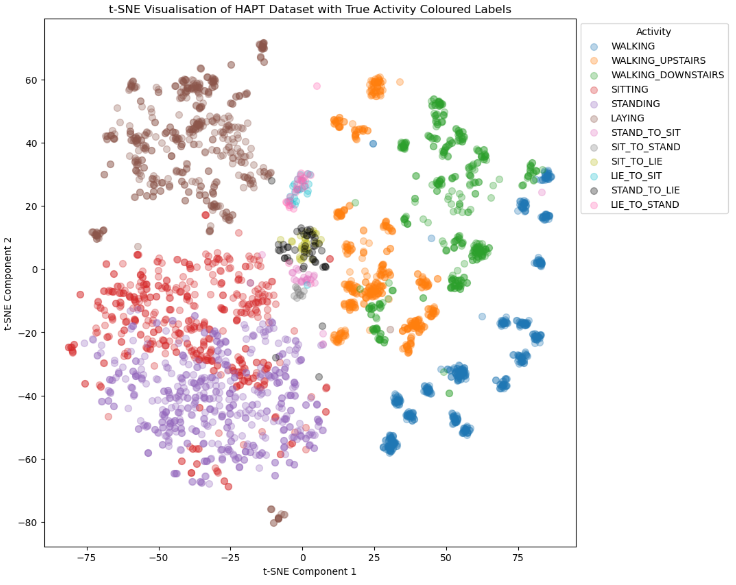


Figure - t-SNE Visualisation of HAPT Dataset with True Activity Coloured Labels

Although most of the clusters and the classes were correctly identified by the RF classifier, the true labels, in the plot above, shows the inaccuracies of the RF classifier. Consider the outlier that was highlighted in the previous plots, the true label shows that this actually belongs to the stand\_to\_sit activity, confirming that this data point is indeed an outlier. Whilst outliers could be removed from the dataset to improve the accuracy of the plots in general, it may be beneficial to preserve the outliers, given the nature of the dataset. As the HAPT data set could be used in healthcare, outliers could highlight areas of interest that may require further investigation, as depending on the purpose of the dataset, outliers may not be dismissible. For this reason, no outliers have been removed from the plots. In addition to the already mentioned outlier, there are numerous other outliers present in the dataset. In order to gain a better understanding of the measurements of each class in the plot, and to better visualise other outliers, an additional twelve plots were created for each class within on the plot, which can be found in the appendices section of this report.

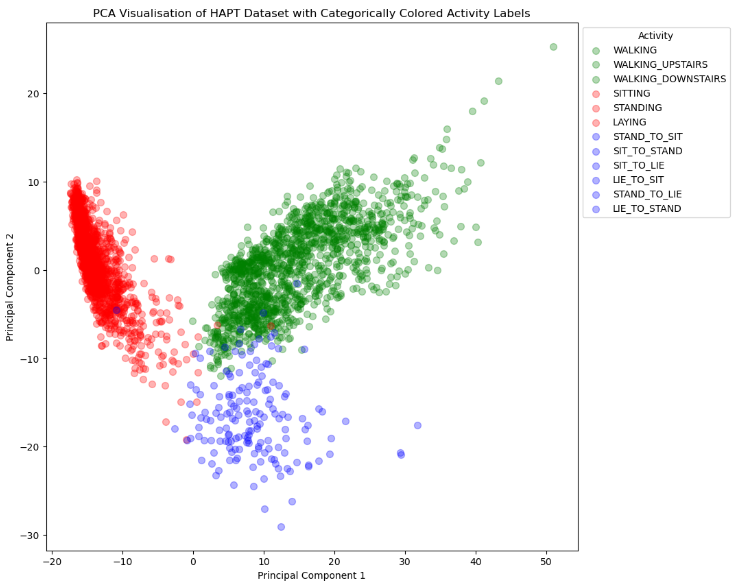
The following plot shows the data points in the reduced feature space that were obtained by applying PCA on the original dataset. This plot presents the same data set that was used in the previous plots, however, a different dimensionality reduction method was used and applied. It is visible that there is some distinction between the 3 activity groups, with some overlaps present from the dynamic movements and the transition movements. However, there are also clearly separated clusters, specifically the tight cluster of the static movements which highlights that there is a potential for a classification algorithm to be utilised. The sections of the plot that show the overlapping data points between all data that could suggest that the PCA failed to capture all the relevant data points. Overall, the plot portrays the dataset with the categorical activities shown in different colours, with somewhat of a divide between the categories. The static measurements have the tightest cluster which indicates that it was mostly successful in differentiating this activity cluster than others.

Figure - PCA Visualisation of HAPT Dataset with Categorically Coloured Activity Labels

The following plot shows the result of the application of PCA to the HAPT dataset along with the RF classifier having predicted the activity classes. Note that this is not the true labels of the data and only a prediction from the classifier with an accuracy score of 55.57%. In regard to classification, 55.57% may not be considered as very high and could indicate that the data is either being overfitted or underfitted, as this can greatly affect the performance of the model.

A screen shot of a diagram

Description automatically generated with low confidenceThe plot above suggests that the RF classifier has been able to make a somewhat decent attempt in predicting the activity labels with getting the categorical activity labels correct, at the very least. In the previous plot, it could be seen that there was an overlap of the different activities and here with the RF classifier, it appears to have classed the majority of the overlap as the “walking\_upstairs” activity. The following confusion matrix was also produced to further evaluate the performance of the RF classifier.

Figure - PCA Visualisation of HAPT Dataset with Random Forest Predicted Labels

A picture containing text, screenshot, font, number

Description automatically generatedThe confusion matrix shows that the RF classifier is not performing too well with the dataset, with the classifier having a large amount of false positive and false negative predictions. Overall, the visualisation shows the PCA-reduced HAPT dataset with the activity labels predicted by the RF classifier, with the main 3 movement categories correctly identified, but the individual activity labels identified mostly incorrectly.

Figure - PCA 2 Confusion Matrix

The following plot is a visualisation of the HAPT dataset with PCA applied but rather than using the predicted activity labels, the true labels are used instead. This will provide a clearer picture of the inaccuracies of the RF classifier as mentioned above.

Figure - PCA Visualisation of HAPT Dataset with True Activity Coloured Labels

A picture containing screenshot, text, colorfulness, diagram

Description automatically generatedAs seen in the plot above, the overlaps between the main three categorical movements mostly consist of the dynamic movements, and the majority of the static movements are clustered together and overlapping. This therefore indicates that such movements contain similarities in their measurements, highlighting the relationships. However, to be able to distinguish between the movements within the main three activity categories may be challenging, due to most of them overlapping, but the three main activity categories themselves can clearly be distinguished. Overall, the visualisation does provide useful information on the relationships between other activity categories, based on their movement patterns. In order to visualise the activity measurements of each individual activity more clearly, additional plots for each activity were also created, which can be found in the appendices section of this report.

Whilst both t-SNE and the PCA plots provided useful plots on how the HAPT dataset can be visualised, the plots created using t-SNE provide a clearer visualisation of the groupings of the different activities and show more similarities in the measurements with less overlaps when compared to the PCA plots.

# Evaluation

This section of the report will discuss the approaches that was used to accomplish the task, including the strengths and weaknesses of the methods used, the choice of the algorithms and techniques, evaluation metrics and highlighting areas for improvement.

The visualisations that were created efficiently provide insight into the HAPT data set and how the measurements of the different activities relate to each other, by highlighting the similarities of the measurements of the activities. Furthermore, clear groupings between the activity categories can be seen within the plots. The RF classifier proved to produce good results when used with t-SNE, suggesting that this machine learning algorithm could be used for future or similar datasets. By using hyperparameter tuning, the effectiveness of the algorithm was significantly improved by finding the best hyperparameters, as apposed to when it was used in testing without the grid search. By using t-SNE and PCA similarly to each other, it was possible to compare the performance of the two methods and therefore, select the method that was best suited for the task. Much like the research carried out by Ramanujam et al. (2021) and Ballı, Sağbaş and Peker (2019), the RF classifier also proved to have good results when applied to the HAPT data set for the task within this report. In addition to the algorithms used, the use of scatter plots was able to accurately present the data visual manner that can be easily understood and interpreted. An experiment was conducted where line charts were used to visualise the data but were found to be a poor representation.

In terms of weaknesses and improvements that could be made, further information regarding the evaluation of the RF classifier could have been included, such as F score, precision and recall. Furthermore, although it was found that the RF appeared to be the best approach for the task, it would have been interesting to see how other classifiers, such as KNN or SVM, would have performed on the task and could have been compared to each other, much like how t-SNE and PCA was. It may have also been beneficial to experiment with the data set itself and see how the results would have differed if a different split between the training and testing data was used. This would have included data for the other 21 participants. It was also noted from the plots that the majority of the measurements that were present were either static or dynamic, as there was significantly less measurements for the transitional movements.

# Summary

In conclusion, this report details the task that was identified and the steps and considerations that were taken in order to complete the goal. The HAPT data set was used to create visualisations by applying machine learning techniques to classify the activities and visualise the data by using t-SNE and PCA. Due to the complex data set, the code used scaling, model training and hyperparameter tuning to create efficient visualisations to provide insights into how the data behaves. The RF classifier was used with grid search cross validation to ensure the best hyperparameters were used to improve accuracy. The plots that were created provides insight into both the true data itself, and how the data can be used with machine learning algorithms, proving that the RF classifier is very competent with the HAPT data. Between the PCA and the t-SNE plots, it was identified that the visualisations created with t-SNE were better used for the insights in the data, due to the information gained by the clusters, points of interest in the outliers, and the higher accuracy score of the RF classifier. Furthermore, this report discussed the purpose of each source code and what was expected to be achieved with each one, with also the findings and understandings of each plot. Finally, the strengths and weaknesses of the of the task were highlighted, also identifying what could have been done differently, or what further work could be done to build from this.

# References

1. Ballı, S., Sağbaş, E. and Peker, M. (2019). Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Measurement and Control -London- Institute of Measurement and Control-*, 52, p.002029401881369. doi:https://doi.org/10.1177/0020294018813692.
2. Hamad, R., Järpe, E. and Lundström, J. (2018). Stability analysis of the t-SNE algorithm for human activity pattern data. doi:https://doi.org/10.1109/SMC.2018.00318.
3. Irfan, S., Anjum, N., Masood, N., Khattak, A.S. and Ramzan, N. (2021). A novel hybrid deep learning model for human activity recognition based on transitional activities. *Sensors*, [online] 21. doi:https://doi.org/10.3390/s21248227.
4. Karim, S., Aqil, S. and Mahmood, N. (2020). An Analysis of Human Activities Recognition using Smartwatches Dataset. *International Journal of Advanced Computer Science and Applications*, 11(12). doi:https://doi.org/10.14569/ijacsa.2020.0111241.
5. Maaten, L. van der and Hinton, G. (2008). Viualizing data using t-SNE. *Journal of Machine Learning Research*, 9, pp.2579–2605.
6. Rahimi Taghanaki, S., Rainbow, M.J. and Etemad, A. (2021). Self-supervised Human Activity Recognition by Learning to Predict Cross-Dimensional Motion. *2021 International Symposium on Wearable Computers*. [online] doi:https://doi.org/10.1145/3460421.3480417.
7. Ramanujam, E., Padmavathi, S., Baskar, L.A. and Niwin, P. (2021). Ensemble Feature Selection for the Recognition of Human Activities and Postural Transitions on Smartphones. *International Journal of Service Science, Management, Engineering, and Technology*, 12(5), pp.80–101. doi:https://doi.org/10.4018/ijssmet.2021090106.
8. Rawat, K. (2020). *Human Activity Recognition based on Energy Efficient Schemes*. MSc Thesis.
9. Shieh, J.-I., Wu, H.-H. and Huang, K.-K. (2010). A DEMATEL method in identifying key success factors of hospital service quality. *Knowledge-Based Systems*, [online] 23, pp.277–282. doi:https://doi.org/10.1016/j.knosys.2010.01.013.
10. Sikder, N. and Abdullah-Al Nahid (2021). KU-HAR: An open dataset for heterogeneous human activity recognition. *Pattern Recognition Letters*, [online] 146, pp.46–54. doi:https://doi.org/10.1016/j.patrec.2021.02.024.
11. Taghanaki, S.R., Rainbow, M. and Etemad, A. (2022). Self-Supervised Human Activity Recognition with Localized Time-Frequency Contrastive Representation Learning. *arXiv:2209.00990 [cs, eess]*. [online] Available at: https://arxiv.org/abs/2209.00990 [Accessed 12 May 2023].

# Appendices

The following visualisations show the individual activity labels plotted alone. This was done to allow all the data points to be completely visualised without any overlaps from other classes. The following plots were created with t-SNE with true labels.

A picture containing screenshot, diagram

Description automatically generated

Figure - t-SNE Visualisation of the Walking Activity with True Labels

A picture containing screenshot, diagram, text

Description automatically generated

Figure - t-SNE Visualisation of the Walking\_Upstairs Activity with True Labels

A picture containing screenshot, diagram

Description automatically generated

Figure - t-SNE Visualisation of the Walking\_Downstairs Activity with True Labels

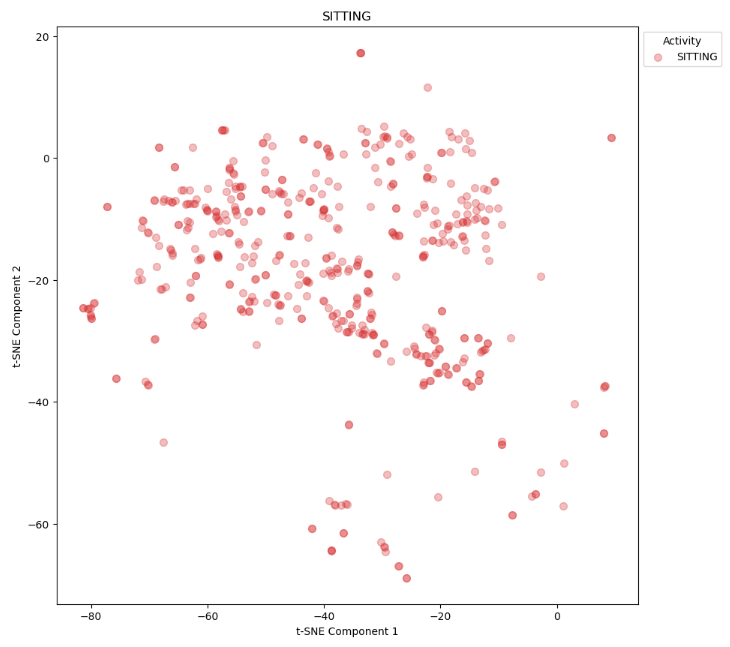


Figure - t-SNE Visualisation of the Sitting Activity with True Labels

A picture containing screenshot

Description automatically generated

Figure - t-SNE Visualisation of the Standing Activity with True Labels

A picture containing screenshot

Description automatically generated

Figure - t-SNE Visualisation of the Laying Activity with True Labels

A screen shot of a graph

Description automatically generated with medium confidence

Figure - t-SNE Visualisation of the Stand\_to\_Sit Activity with True Labels

A screen shot of a graph

Description automatically generated with low confidence

Figure - t-SNE Visualisation of the Sit\_to\_Stand Activity with True Labels

A screen shot of a graph

Description automatically generated with low confidence

Figure - t-SNE Visualisation of the Sit\_to\_Lie Activity with True Labels

A screen shot of a graph

Description automatically generated with medium confidence

Figure - t-SNE Visualisation of the Lie\_to\_Sit Activity with True Labels

A screen shot of a graph

Description automatically generated with low confidence

Figure - t-SNE Visualisation of the Stand\_to\_Lie Activity with True Labels

A screen shot of a graph

Description automatically generated with medium confidence

Figure - t-SNE Visualisation of the Lie\_to\_Stand Activity with True Labels

The following plots were created with PCA with true labels.

A picture containing screenshot, diagram, text

Description automatically generated

Figure - PCA Visualisation of the Walking Activity with True Labels

A picture containing screenshot

Description automatically generated

Figure - PCA Visualisation of the Walking\_Upstairs Activity with True Labels

A picture containing screenshot, diagram

Description automatically generated

Figure - PCA Visualisation of the Walking\_Downstairs Activity with True Labels

A picture containing screenshot

Description automatically generated

Figure - PCA Visualisation of the Sitting Activity with True Labels

A picture containing screenshot

Description automatically generated

Figure - PCA Visualisation of the Standing Activity with True Labels

A picture containing screenshot

Description automatically generated

Figure - PCA Visualisation of the Laying Activity with True Labels

A screen shot of a graph

Description automatically generated with medium confidence

Figure - PCA Visualisation of the Stand\_to\_Sit Activity with True Labels

A screen shot of a graph

Description automatically generated with medium confidence

Figure - PCA Visualisation of the Sit\_to\_Stand Activity with True Labels

A picture containing text, screenshot, rectangle, line

Description automatically generated

Figure - PCA Visualisation of the Sit\_to\_Lie Activity with True Labels

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

Figure - PCA Visualisation of the Lie\_to\_Sit Activity with True Labels

A picture containing diagram, screenshot, text

Description automatically generated

Figure - PCA Visualisation of the Stand\_to\_Lie Activity with True Labels

A picture containing text, screenshot, diagram

Description automatically generated

Figure - PCA Visualisation of the Lie\_to\_Stand Activity with True Labels