# FEATURE REDUCTION USING PCA, LDA, AND PCA+LDA TO IMPROVE DECISION TREE C4.5 CLASSIFICATION OF HUMAN ACTIVITY RECOGNITION DATASET

**Matthew Cann**[1]

1: Department of Mechanical Engineering, University of Waterloo, Waterloo, ON, Canada, mcann@uwaterloo.ca

## Abstract

In this study, three methods of feature reduction were analyzed to investigate the effect on accuracy and computational cost of a C4.5 classifying decision tree algorithm on the Human Activity Recognition (HAR) Using Smartphones dataset from the UCI Machine Learning Repository. The dataset is comprised of 561 attributes measured from the accelerometer and gyroscope of a smartphone over 10299 instances. Feature reduction on the dataset helps to solve the overfitting behaviour of the C4.5 algorithm when fitting data with many extraneous features. The dimension reduction techniques analysed were principal component analysis (PCA), linear discriminant analysis (LDA), and a PCA+LDA combination. The feature reduction methods were applied to the training set to transform the data. Using the reduced dataset, a C4.5 decision tree was trained and then used to predict the class labels using the transformed testing dataset. From the analysis of this study, LDA outperforms the other dimension reduction methods in accuracy, precision and computational complexity resulting in a final testing accuracy on the HAR dataset of 95.62%.

## 1 Summary of Feature Reduction

The goal of feature extraction in classifying applications is to reduce the feature space of the original data while still representing the data and class label relationship. Reduced feature space is desirable in classification to reduce features that add noise to the solution and inter-feature correlations [1] which addresses the overfitting problem of the C4.5 algorithm.

PCA feature extraction was used by Nasution [2] for a C4.5 classification tree and showed an increase in accuracy of 4.65% compared to no feature reduction. The feature space was reduced from 36 attributes to 12 by considering the eigenvalues that comprised of 99.80% of the information. The author reported an increase in particularity and precision of the results. Varmaa et al. [3] reduced the feature space by half using PCA alongside a modified Fuzzy SLIQ decision tree (FSDT) algorithm. The authors achieved the highest accuracy compared to multiple publications using the same PID dataset from the UCI machine learning repository. The combination of PCA feature extraction and FSDT algorithm achieved an improvement of accuracy of 1.8% compared to the C4.5 algorithm. Neither of the publications mentioned the effect of the lower feature space on the computational resources required to perform the classification steps.

PCA and LDA methods are common for feature extraction of signals for classification using Support Vector Machine (SVM) [4] [5]. Subasi [4] classified epileptic seizures using EEG signals and an SVM classifier. The accuracy of classification was increased to 98.75% and 100% using the dimensional reduction methods PCA and LDA, respectively. Leo et al. [5] investigated an efficient approach for preprocessing data from a chemical sensor array which included PCA as well as LDA for a Support Vector Machine (SVM) classifier. The author demonstrated the ability to use LDA to better separate the data to be classified resulting in higher accuracy. The author extends the quantitative metrics to include elapsed time to compare the computational complexity of the methods. The effects of PCA and LDA on the fitting and predicting computational resources are not thoroughly explored in this work. Pechenizkiy [6] utilized PCA and LDA feature extraction on a variety of machine learning models and proposed a novel feature extraction method which aimed to overcome the deficiencies of the methods by combing them. By combining the extraction linear discriminants with a few principal components, the accuracy of the classifiers improved over many datasets. The largest improvement was with C4.5 decision trees, but the specific improvements depend on the nature of the dataset. Over the 21 datasets that the analysis was conducted, the average accuracy of LDA was slightly larger (0.754) compared to PCA (0.746) for C4.5 decision trees. The combination of LDA+PCA showed a 3.3% increase in accuracy compared to PCA and a 2.5% increase compared to LDA results. Yang [7] provides the background and proof of using LDA on PCA transformed space and recommends using more PCA dimensions such that the LDA algorithm can perform better. Li [8] utilized the PCA and LDA combination feature extraction for face recognition using the nearest neighbour classifier which showed improved results compared to PCA alone. The largest improvement in recognition rate (accuracy) was reported to be 9.4% between the PCA and PCA+LDA using 40 feature dimensions. The author demonstrated a diminishing return of accuracy as the feature vector dimension increased. This may be a consequence of the peaking phenomena of PCA for the return number of dimensions.

This paper aims to analyzes the effect of PCA, LDA, and PCA+LDA on the accuracy and precision of the C4.5 classifying decision tree algorithm. Furthermore, the computational resource effect of the varying feature reduction methods was analysed through the computational time required for the model.

# 2 Justification and Methodology

The C4.5 decision tree algorithm uses gain ratio as the splitting criteria extending from the original ID3 algorithm. The C4.5 algorithm tends to overfit data that has many irrelevant features causing an increase in misclassification on testing sets [2]. Feature reduction is a solution to reduce overfitting by reducing the dimensions of the training and testing data. Feature reduction was used in this analysis because of the large number of features totalling 561. By performing a feature reduction, the most important features were extracted to reduce the number of extraneous features and improve the classification ability of the model.

The framework of this study is to perform PCA, LDA, and PCA+LDA feature reduction methods on the training and testing data as a preprocessing step. The next step was to train a classification decision tree using the C4.5 algorithm with the transformed data to improve the accuracy of the model. The objective of the study is to determine the optimal feature reduction method for the Human Activity Recognition Using Smartphones Data Set. The objective was completed through the following phases.

- Data Acquisition: The Human Activity Recognition Using Smartphones Data Set was used from the UCI Machine Learning Repository.
- Data standardization: Standardizing the data is an important step to ensure that the different parameters and scales of the features are transformed into a similar scale. The data from the UCI Machine Learning Repository are already normalized and bounded within [-1, 1]
- Data Preprocessing: The training and testing data was transformed using PCA, LDA, and the combination of PCA and LDA.
- Build a classification model using the C4.5 decision tree: The model was trained using the transformed dataset for each method.
- Prediction: Using the fitted model of each technique, the test dataset is used to predict

the class label. To evaluate the performance of the classifier with the different feature reduction methods the accuracy and standard deviation are reported. The effect of the pre-processing stage on the computational cost of the model is evaluated using the processing time required in the fitting stage. Furthermore, a confusion matrix of the classification was reported to gain understanding the types of misclassification each trial exhibits.

## 2.1 Dataset Description

The UCI Machine Learning Repository was used for the Human Activity Recognition (HAR) Using Smartphones Data Set. The dataset has 6 class labels to represent the following: 1 WALKING, 2 WALKING_UPSTAIRS,3 WALKING_DOWNSTAIRS, 4 SITTING, 5 STANDING, 6 LAYING. A total of 561 features were recorded from a waist-mounted smartphone with inertial sensors for a total number of instances of 10299. The features are a combination of triaxial acceleration from the accelerometer in the smartphone and the estimated body acceleration. From the gyroscope, the triaxial angular velocity values were recorded.

## 2.2 Principal Component Analysis (PCA)

The principal component analysis is an unsupervised dimensional reduction method that determines the most accurate data representation in a lower-dimensional space while preserving the largest variance in the data. In this study, PCA was applied to both the training and testing set of the HAR data. The principal component analysis was conducted in Python using PCA decomposition function in the Scikit Learn library (v0.22.2) [9]. The peaking phenomena of the PCA were investigated to achieve the optimal number of feature dimensions to be used in the analysis. The accuracy of the C4.5 decision tree with varying feature dimensions of PCA is shown in Figure 1.

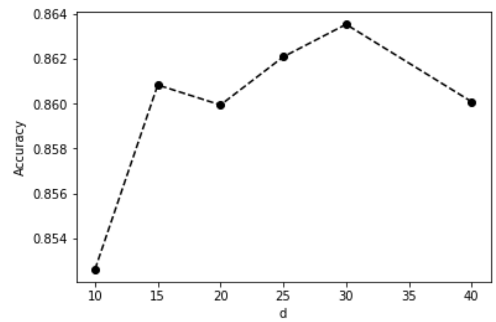



Figure 1. PCA C4.5 accuracy with varying feature dimension.

The optimal number of dimensions for the C4.5+PCA case was determined to be 30 from the peak of Figure 1, this number of dimensions is carried through for further analysis.

## 2.3 Linear Discriminate Analysis (LDA)

Linear discriminant analysis performs dimensionally reduction while preserving the largest variance in the class information. LDA aims to maximize the distance between the projected class means while keeping small variance within each class. LDA was conducted to transform both the training and testing set of the HAR dataset and was trained using both the X training set and y training class labels. The linear discriminate analysis was conducted in Python using LinearDiscriminantAnalysis function in the Scikit Learn library (v0.22.2) [10]. The feature space of LDA is limited to the number of class labels minus one, in this case, all five features were used in the analysis.

## 2.4 PCA+LDA Combination

The LDA performed in the PCA space was achieved by first transforming the training and testing data using PCA. The second feature extraction is performed on the transformed training and testing data using LDA. Following the method of Yang [7], the PCA+LDA combined transformation is performed for a varying number of PCA dimensions. The training and testing data are then inputted into the C4.5 classifying decision tree to analysis the peaking phenomena as shown in Figure 2.

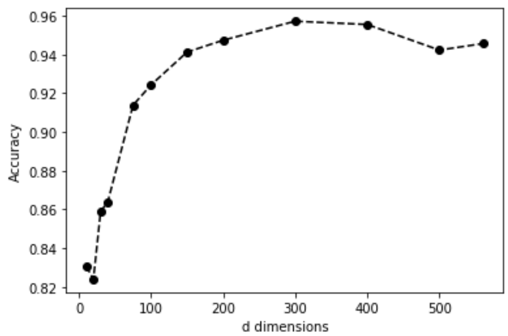



Figure 2. PCA+LDA C4.5 accuracy with varying feature dimension.

The optimal number of dimensions for the PCA+LDA combination was determined to be 300, shown by the maximum occurring in Figure 2.

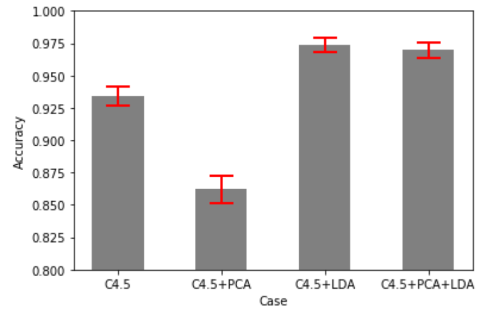
# 3 Results and Discussion

The four cases of varying feature extraction methods (None, PCA, LDA, PCA+LDA) were used to train a C4.5 classifying decision tree. The trained model was used to classify the testing set and the accuracy was reported based on 10-times 10-fold cross-validation. The computational cost of the techniques was evaluated by reporting the processing time to train the model. The training time was chosen to be reported since it was determined to be the most substantial time. All the techniques were implemented using Python 3.7.3 on a system with Intel Core i5-8250U CPU @1.60GHz and 8.00GB of RAM. The results of the accuracy and training time with the corresponding standard deviation (STD) values for the varying techniques are summarized in Table I.

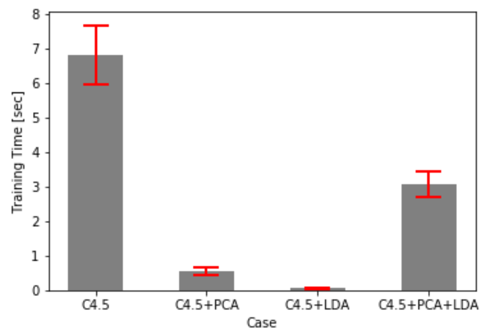
TABLE I: 10-TIME-10-FOLD CROSS VALIDATION ACCURACY & TRAINING TIME RESULTS

| | **Accuracy [%]** | **STD [%]** | **Train Time [sec]** | **STD [sec]** |
|---|---|---|---|---|
| C4.5 | 93.53 | 0.8659 | 6.814 | 0.8612 |
| C4.5+PCA | 86.12 | 1.106 | 0.5544 | 0.1091 |
| C4.5+LDA | 97.37 | 0.5224 | 0.05903 | 0.01709 |
| C4.5+LDA +PCA | 96.97 | 0.6055 | 3.062 | 0.3752 |

The varying results of accuracy and training time are visualized using bar graphs as shown in Figure 3.




(a)

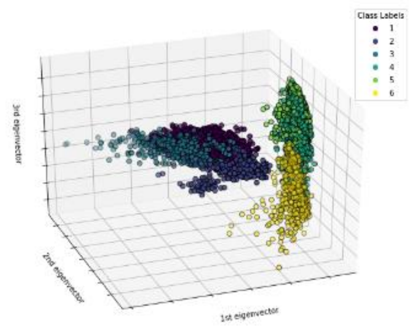



(b)

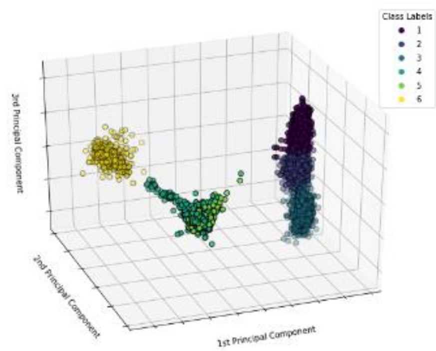
Figure 3. Bar graph of the results of feature extraction techniques on (a) accuracy and (b) training time.

From the results of the cross-validation, using LDA in the preprocessing step of the C4.5 classifying tree was determined to be the best setup studied in this analysis. The C4.5 decision tree accuracy for classifying was improved by 4.1% while decreasing the computational complexity to train the model by 99% for the HAR dataset. The precision of the model is improved using LDA by 40% as shown by the lower standard deviation in the accuracy results.

The poor results of the PCA preprocessing can be attributed to the class labels in the reduced space not being separated as efficiently compared to the LDA method. The classes of the data for the first three features are shown in Figure 4.




(a)




(b)

Figure 4. First 3 features from (a) PCA and (b) LDA feature reduction.

The effectiveness of LDA for this dataset is apparent by the separated classes in the feature space of the transformed data in Figure 4. The difference in the transformed data is attributes to how PCA will weight features with higher variance despite the class label variance due to the unsupervised nature.

PCA, in this case, provides uncorrelated linear combinations of the dataset and works the best for independent features. The HAR dataset has spatial features that are not independent which leads to the ineffective results of the PCA. Furthermore, PCA determines the features with the greatest variance which does not always imply that these features are the most important features for classification. LDA addresses the problem of PCA by using class information to determine the best features that represent the classes. The limitation of LDA is the maximum number of features that can be extracted is one minus the number of class labels. In our case, the 5 features determined from LDA had enough class separation to provide a good classifier.

Using LDA for feature reduction alongside the C4.5 algorithm, the final accuracy of the training and testing data from the UCI Machine Learning Repository was determined. The training and testing data from the repository are approximately a 70/30 split, respectively. The test accuracy of the best model setup determined by this study was 95.62% with the confusion matrix shown in Figure 5.

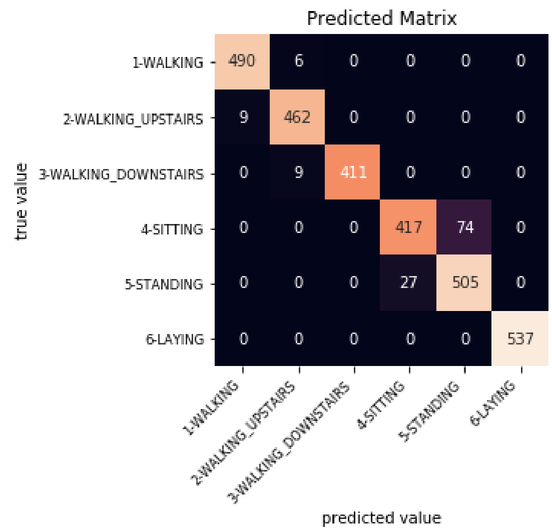



Figure 5. Confusion matrix for the test accuracy of 95.62%.

The testing accuracy achieved in this study outperforms previous work with the HAR dataset using C4.5 by 1.61% [11].

# 4 Conclusion and Recommendations

In this study, the effects of dimensionality reduction were investigated on the results of a C4.5 classifying decision tree algorithm on the Human Activity Recognition (HAR) Using Smartphones dataset from the UCI Machine Learning Repository. Feature reduction was used to solve the trend of the C4.5 algorithm to overfit data that has many irrelevant features that reduce the test accuracy. The dimension reduction techniques analysed were principal component analysis (PCA), linear discriminant analysis (LDA), and a PCA+LDA combination. The feature reduction methods were applied to the training set to transform the data. Using the reduced dataset, a C4.5 decision tree was trained and then used to predict the class labels using the transformed testing dataset. The results of this study show that LDA outperforms the other dimension reduction methods with increased accuracy of 4.1% using 10-time-10-fold cross-validation, improving the precision of the model by 40% while decreasing the computational complexity to train the model by 99%. The final testing accuracy on the HAR dataset was 95.62% which outperforms previous work using the same dataset with the C4.5 algorithm by 1.61% [11].

The results could be improved by extending the number of features beyond accelerometer and gyroscope signals. The classification results may benefit from data acquired from the smartphones barometer, heart rate monitor, Bluetooth, GPS, and microphone. The methods in this study could be combined with other feature selection methods such as filter or wrapper based.  Other forms of feature reduction and selection could be investigated such as a random forest. This method uses random forest classifiers that are trained using the data, then the most important features are identified using a threshold of importance. These features are then used to create a new subset of data which can be used to train a classifier. The results of this study could be extended by utilizing non-linear PCA strategies such as Kernal methods. Furthermore, this study only used a single decision tree classifier, the results could be improved using Random Forest classifiers.

# 5 References

[1] T. K. Oza N.C., "Dimensionality Reduction Through Classifier Ensembles," Technical Report NASA-ARC-IC-1999-124, Computational Sciences Division, NASA, Moffett Field, CA, 1999.

[2] M. Z. Nasution, O. S. Sitompul and M. Ramli, "PCA based feature reduction to improve the accuracy of," *Journal of Physics: Conf. Series,* vol. 978, pp. 1-6, 2018.

[3] K. Varmaa, A. A. Rao, S. T. Mahalakshmi and N. P. Raoa, "A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach," *Applied Soft Computing,* vol. 49, pp. 137-145, 2016.

[4] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications,* vol. 37, p. 8659–8666, 2010.

[5] M. Leo, C. Distante, M. Bernabei and K. Persaud, "An Efficient Approach for Preprocessing Data from a Large-Scale Chemical Sensor Array," *Sensors,* vol. 14, pp. 17786-17806, 2014.

[6] M. Pechenizkiy, A. Tsymbal and S. Puuronen, " On combining principal components with Fisher's," *Foundations of Computing and Decision Sciences,* vol. 31, no. 1, pp. 59-74, 2006.

[7] J. Yang and J.-y. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognition,* vol. 36, p. 563 – 566, 2003.

[8] J. Li , B. Zhao and H. Zhang, "Face Recognition Based on PCA and LDA Combination Feature Extraction," in *First International Conference on Information Science and Engineering,* Nanjing, 2009.

[9] "sklearn.decomposition.PCA — scikit-learn 0.22.2 documentation," Scikit-learn.org, 2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html. [Accessed 14 04 2020].

[10]"sklearn.discriminant_analysis.LinearDiscriminantAnalysis — scikit-learn 0.22.2 documentation," Scikit-learn.org, 2019. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html. [Accessed 20 04 2020].

[11] M. Luštrek and B. Kaluža, "Fall Detection and Activity Recognition with Machine

Learning," *Informatica,* vol. 33, p. 205–212, 2009.

[12] A. Naidu, "Feature Extraction," 2020. [Online]. Available: https://github.com/ashushekar/feature-extraction. [Accessed 14 04 2020].