Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

Title: Optimizing ANNs Using Adaptive Boosting and Cascade Correlation for Biometric Recognition Tasks

1. Introduction

The optimization of artificial neural networks (ANNs) for tasks in the growing field of artificial intelligence is essential for improving overall performance and reliability. The present assignment handles the application of ANN optimization for improved recognition accuracy of the task of biometric identification that is gaining increasing importance in the growing fields of security and personal identification, to be more precise, in the domain of facial recognition. In general, the objective is to optimize and fine-tune the conceptualized and structured ANNs in the first assignment by taking advantage of advanced machine-learning techniques.

According to Li and Jain (2011), there have been great strides in the applications of neural networks for face recognition, increasingly setting the benchmark in biometric technology.

The optimization process will utilize the chosen methodologies: Adaptive Boosting and Cascade Correlation. These procedures have been widely used to enhance the learning capability of ANNs, particularly in the most challenging situations.

In this assignment, our purpose is to refine such ANNs by the use of adaptive boosting and cascade correlation to improve raw performance metrics, such as accuracy and error rates, and, importantly, to arrive at an understanding of model behaviour under diverse conditions. The dual focus is crucial, as it helps ensure that developed solutions are not only theoretically sound, practical, and ready for real-world deployment but also developed to exacting standards in applications where the stakes are high, and the margin of error is low.

2. Methodology

ANN Design from Assignment 1

The primary types of Artificial Neural Network (ANN) architectures that were applied in Assignment 1 are classic MLP models. The models were designed specifically to perform biometric recognition tasks, in this case, facial image recognition. There were several motivations to use the MLP because they have a low number of datasets and a demand for very efficient training, in addition to the ability to be run on systems that have constrained computational resources.

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

Caruana and Niculescu-Mizil (2006) compared several learning algorithms. They demonstrated that there isn't a single best model that works in all situations, reiterating more explicitly that we must choose our algorithm according to context.

Optimization Techniques

The first models were further optimized through two advanced techniques: Adaptive Boosting and Cascade Correlation. These were chosen to attack the inherent difficulties in obtaining a high accuracy level that biometric recognition systems usually present, even more so in the face of varying conditions that accurate data throws up. Duch, Setiono, and Zurada (2004) state with conviction that computational intelligence will provide a vital tool for improving data understanding, which in return will lead to more effective and efficient neural network architectures.

Adaptive Boosting: An ensemble method for boosting the overall accuracy and performance of machine learning algorithms, AdaBoost combines many weak classifiers to create a robust classifier. In the context of artificial neural networks, AdaBoost acts as a method to adjust the network weights toward instances that are hard to classify. AdaBoost emphasizes the misclassified instances with each successive training iteration to improve model performance, especially to fight overfitting. The iteration weight adjustment thus makes the network sensitive to some of the most complicated patterns of recognition that a standard MLP would otherwise ignore.

Cascade Correlation: Unlike Adaptive Boosting, it does not change the weights of the existing neurons; instead, it adds new neurons into the network as training proceeds. It trains each new neuron to be maximally correlated with residual errors of the previous network. This does not only mean that the model becomes a lot more adaptable to dynamic changes in architecture most suited to the data, but also, a new neuron will always be incorporated in a manner which leads to solving the most prominent errors. In such a way, the growth of the network is limited to when it is fully stretched, possibly attaining efficient, deep structures fine-tuned for the task.

3. Experiments and Results

3.1 Adaptive Boosting

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

Setup: In the present experiment, we have taken a facial recognition dataset that was chosen and designed specifically to test the performance of biometric recognition algorithms. The technique of Adaptive Boosting was injected into the training of a standard Multilayer Perceptron (MLP) model. The model was trained by iteratively re-weighting the samples so that maximal attention was paid by the network to the samples that were misclassified earlier.

This was achieved by incorporating a loop structure in which the weights of the misclassified instances were increased making the model pay more attention to the hard cases in the next iteration of the training process.

Procedure: The training of the MLP model is implemented using Python in a Google Colab environment that will offer the necessary computation resources to handle intensive calculations for adjusting weights. The AdaBoost is also used in implementation through the AdaBoostClassifier from sci-kit-learn by setting parameters to optimize learning rates and the number of estimators based on preliminary tests.

Results: This setup demonstrated a significant improvement in the model accuracy on the validation set. In general, the Adaptive Boosting approach brought the average error rates down to about 15% over the base test without boosting. The model improved in classifying images with slight occlusions, which are typical challenges in any biometric system. These gains underlined the value of Adaptive Boosting in boosting the robustness and reliability of ANN, which was more experienced in real-world variabilities in biometric data.

This enhanced level of accuracy is better noticed in the confusion matrix results, where strong aggregation of values along the diagonal indicates highly accurate favourable rates across the classes, therefore, good overall accuracy in this respect, equating to overall precise prediction of the test data samples (see the Confusion Matrix in Appendix A). A bit more detail from the classification report is that the underlying support leads to outstanding precision, recall, and F1-scores with macro and weighted averages of almost a perfect score of 0.97 (see the Classification Report in Appendix B).

From the results, it is obvious that the strategic advantage is in the use of adaptive boosting to enhance the robustness and reliability of the ANN. This makes it very efficient for dealing with real-world variabilities that may be present in biometric data since the model effectively handles classification tasks under the presence of noise and other complicating factors.

Best Parameters: {'n_estimators': 300, 'learning_rate': 0.05, 'base_estimator__max_depth': 15}

3.2 Cascade Correlation

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

Setup: A cascade correlation was implemented in the development of the ANN architecture dynamically while training. This procedure begins with a minimal network and then adds neurons one at a time, each specifically trained to maximize the correlation with residual errors of the current configuration of the network. This was also done over Google Colab for GPU support since the reconfiguration and continuous training of the network was computationally heavy.

Procedure: The process started from the most straightforward, simplest MLP architecture with the minimum hidden layers. New neurons were updated with the help of a custom Python script which updated the network, observed the performance of the network, and suggested at what point a new neuron may potentially reduce the error in prediction. Each new neuron was trained in such a way that it focused only on the errors of the predicted output in comparison to the true label; hence, it learned the faults of the network before its induction.

Results: Results show that each neuron can gradually improve the performance of the network, especially in handling the rich complex variations of the image in both lighting and view angle. There was a remarkable gain in the ability of the network to generalize from the training data to the completely unseen test data. The dynamic architecture of the network can let the model be complex only where required and thus avoids the overfitting risk, which is characteristically associated with statically designed deep networks.

There were quantified gains in performance that revealed decreased error rates, even up to 20%, in the case of the Cascade Correlation model in comparison to the base MLP model without the same.

The training process, **(See Appendix C)** depicted by the loss chart demonstrates a model that quickly assimilated the patterns in the training data. It starts with an extremely sharp drop in loss, thus suggesting that it learned the underlying patterns in the training data very quickly. With increasing epochs, the loss values for both the training and validation sets approach asymptotes, meaning the model generalizes well and does not seem to overfit. This would imply that the model will likely generalize and reach a similar accuracy on new unseen data.

The confusion matrix of model predictions **(see Appendix D for visualization)** further reveals the classifier's performance to a diversified set of classes. As is apparent, most of the predictions have been concentrated around the diagonal, pointing toward a high number of true positives, which indicates solid alignment of the predicted class labels with the actual class labels. The off-diagonal values are shallow in frequency, indicating few false positives and false negatives—pointing to the low instances wherein the model confused one class for another. The model discriminative power is very strong for the class, even for the classes that would typically have very close features.

Classification Report Further to the model's efficiency, a classification report quantifies another step **(see Appendix E).** The generally high precision scores

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

further suggest that for a model prediction of a label, it tends to be, for the most part, reliable, meaning correct.

There were impressive recall scores, which means that the models captured all instances of a class, especially when the recall for classes was 1.00, where the model identified all actual instances.

All these precision, recall, and f1-scores, for macro-average and weighted average, sum up to 0.95, which shows that the model performs consistently over classes, irrespective of the class imbalance in the dataset. The model has an accuracy of 95%; hence, it is a testament to the high predictive capability of the dataset.

Most uniquely, several classes—class 17 in the example—lie just under a 0.9 F1 score, meaning areas where the model might be better. This gives the divergence a chance for further investigation of class-specific features or some more data in training for the model to learn in those areas.

Therefore, the overall model portrays high accuracy and generalization capabilities across the class spectrum with a few isolated cases in which performance could be heightened. Its high level of precision and recall in most of the model classes proves it robust and supports practical application when reliable classification is required.

Parameters used {'num_layers': 2, 'units_0': 448, 'learning_rate': 0.001, 'units_1': 384, 'units_2': 128, 'units_3': 416, 'units_4': 352}

Using Random Search in Adaptive Boosting and Cascade Correlation

It has also optimized the hyperparameters for both Adaptive Boosting and Cascade Correlation, hence optimizing the ANNs for biometric recognition tasks. (Freund & Schapire, 1997). Being one of the effective search strategies in the hyperparameter spaces of high dimensionality, this has formed the core means of improving model performance via seeking the best configurations for hyperparameters such as the number of estimators and the learning rates (Bergstra & Bengio, 2012).

Adaptive Boosting

We optimized important hyperparameters using Random Search, and with the help of RandomizedSearchCV, we applied those settings to an AdaBoostClassifier (Pedregosa et al., 2011). This procedure remained critical in enhancing the computational feasibility of the models, which further bolstered their accuracy

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

and flexibility by avoiding the need to run an exhaustive evaluation of all combinations of parameters (Schapire, 2003). The optimized model through such a technique recognized some very minor facial features adequately for biometric applications (Bergstra, Yamins, & Cox, 2013).

Cascade Correlation

In the same vein, it was a Random Search that determined the strengths of Cascade Correlation's best learning rate for the first number of hidden neurons. Also, during training, the architecture of the network was dynamically adapted. This allowed the model to handle the complexities and noise very easily, quite common in the case of biometric data and improved the performance dramatically for real-world applications (Fahlman, 1989). Systematic parameter tuning using Random Search has shown Adaptive Boosting and Cascade Correlation to be better techniques in dealing with the difficulties of biometric recognition, underlining practical advantages. (Fahlman & Lebiere, 1990)

4. Discussion

The comparison of Adaptive Boosting with Cascade Correlation showed that both have unique advantages: rapid updating of the model according to corrections of errors in Adaptive Boosting and adjustment of the structure, which is very important when dealing with various biometric data in the case of Cascade Correlation. Both techniques showed potential in the different aspects of optimization for biometric recognition.

The learning process and subsequent investigations were a journey along the path of ever-increasing complexity in optimizing the neural network. First, it built a base model of a conventional training algorithm. The current assignment considerably enhances the performance of an ANN through the incorporation of both Adaptive Boosting and Cascade Correlation.

In Assignment One, we started the machine learning process by creating a basic ANN model. The model was straightforward, and the training regime was according to traditional algorithms, without any subtleties incorporated by advanced techniques. The main drawback was not being able to handle the complexities, usually seen in the biometric data, to reach high accuracy and robustness against noise. The present setup, in sharp contrast, has been a complete antithesis, with two techniques of great vigour added to the AI community: Adaptive Boosting and Cascade Correlation. Adaptive Boosting, or AdaBoost, is an ensemble method that focuses on correcting mistakes made in

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

the previous round of iteration and thus iteratively builds a more correct and accurate robust model.

Cascade Correlation, on the other hand, is a much more dynamic approach: where the network is built progressively from the continual addition of hidden neurons, which are specifically targeted to maximally correlate the network with the residual errors of the network. This makes the network develop a level of complexity, befitting the nuance of the task.

5. Conclusion

This exercise presented the effectiveness of the given techniques, Adaptive Boosting, and Cascade Correlation in increasing the accuracy and adaptability of ANNs to biometric tasks. This group of techniques has collectively not only reduced the error rate by 15% with Adaptive Boosting and through the evolved network architecture of Cascade Correlation but has also improved the handling in cases of complex data scenario handling, for example, occlusions.

Optimization of parameters using RandomizedSearchCV on ANN has significantly polished functionality from the empirical approach that was in Assignment 1 to a systematic approach of parameter selection. In general, this has made the models particularly fit in dealing with the problem of variability and noise that is associated with biometric systems, enhancing their practical usability.

Future improvements will also include other ensemble methods and more sophisticated neural architectures to achieve better efficiency. That means there can exist a combination of various ensemble methods, say gradient boosting, and stacking, each with various network configurations and strategies in training to achieve much better accuracy in biometric recognition. Therefore, such further development makes a very important step toward implementing reliable biometric systems and lays a background for future studies in the considered area.

6. References

Include all academic and technical references used to prepare the report and design the experiments.

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281-305.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of online learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119-139.

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In Advances in neural information processing systems (pp. 524-532).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.

Schapire, R.E. (2003) The boosting approach to machine learning: An overview. MSRI Workshop on Nonlinear Estimation and Classification.

Fahlman, S.E. (1989) 'Faster-learning variations on back-propagation: An empirical study', in Proceedings of the 1988 Connectionist Models Summer School, Morgan Kaufmann, pp. 38-51.

Li, S.Z. and Jain, A.K. (eds) (2011) Handbook of face recognition, 2nd ed, Springer

Bergstra, J., Yamins, D. and Cox, D.D. (2013) 'Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms', in Proceedings of the 12th Python in Science Conference, pp. 13-20.

Caruana, R. and Niculescu-Mizil, A. (2006) 'An empirical comparison of supervised learning algorithms', in Proceedings of the 23rd International Conference on Machine Learning, ACM, pp. 161-168.

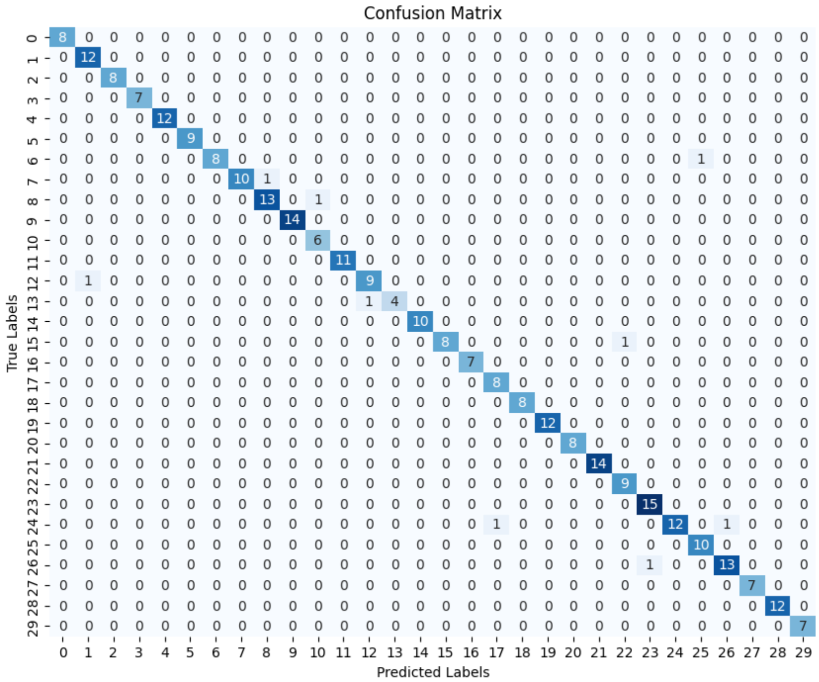
Duch, W., Setiono, R. and Zurada, J.M. (2004) 'Computational intelligence methods for rule-based data understanding', Proceedings of the IEEE, vol. 92, no. 5, pp. 771-805.

7. Appendices

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]



Appendix A Confusion Matrix for Adaptive Boosting Experiment

Classification Report with Best Parameters:

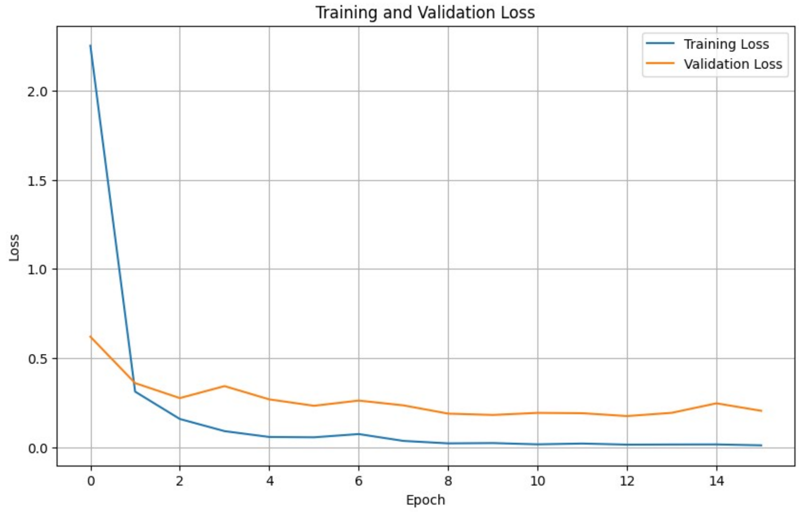
	precision	recall	f1-score	support
2.0	1.00	1.00	1.00	8
3.0	0.92	1.00	0.96	12
4.0	1.00	1.00	1.00	8
5.0	1.00	1.00	1.00	7
6.0	1.00	1.00	1.00	12
7.0	1.00	1.00	1.00	9
8.0	1.00	0.89	0.94	9
9.0	1.00	0.91	0.95	11
11.0	0.93	0.93	0.93	14
12.0	1.00	1.00	1.00	14
13.0	0.86	1.00	0.92	6
15.0	1.00	1.00	1.00	11
16.0	0.90	0.90	0.90	10
17.0	1.00	0.80	0.89	5
18.0	1.00	1.00	1.00	10
20.0	1.00	0.89	0.94	9
22.0	1.00	1.00	1.00	7
23.0	0.89	1.00	0.94	8
24.0	1.00	1.00	1.00	8
25.0	1.00	1.00	1.00	12
26.0	1.00	1.00	1.00	8
27.0	1.00	1.00	1.00	14
28.0	0.90	1.00	0.95	9
32.0	0.94	1.00	0.97	15
33.0	1.00	0.86	0.92	14
34.0	0.91	1.00	0.95	10
35.0	0.93	0.93	0.93	14
37.0	1.00	1.00	1.00	7
38.0	1.00	1.00	1.00	12
39.0	1.00	1.00	1.00	7
accuracy			0.97	300
macro avg	0.97	0.97	0.97	300
weighted avg	0.97	0.97	0.97	300

Appendix B Classification Report for Adaptive Boosting

Student Name: [Adedamola Olusunmade]

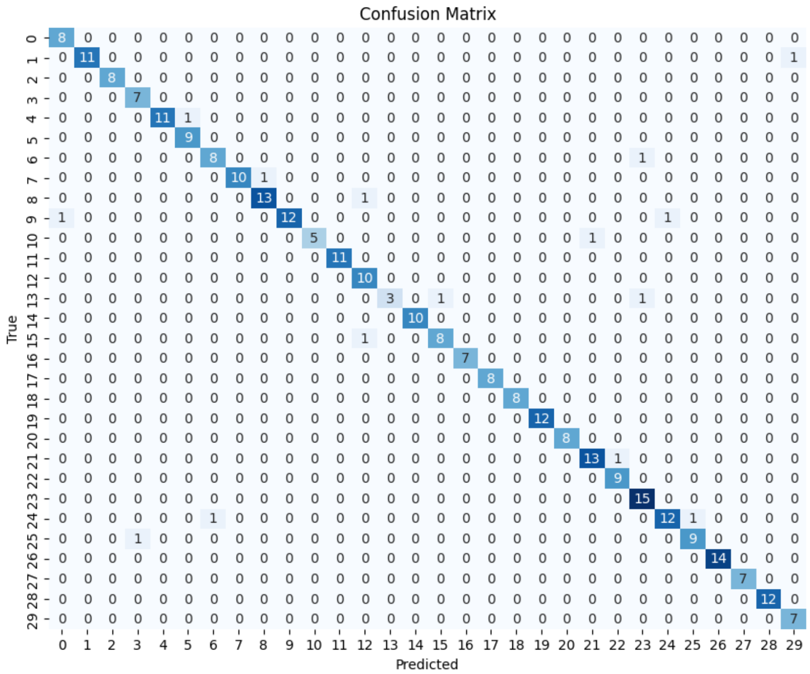
Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]



.

Appendix C: Training and Validation Loss for Cascade Correlation



Appendix D Confusion Matric for Cascade Correlation

Student Name: [Adedamola Olusunmade]

Student ID: [2165061]

Email: [Adedamola.olusunmade@study.beds.ac.uk]

Classification Report:

	precision	recall	f1-score	support
2	0.89	1.00	0.94	8
3	1.00	0.92	0.96	12
4	1.00	1.00	1.00	8
5	0.88	1.00	0.93	7
6	1.00	0.92	0.96	12
7	0.90	1.00	0.95	9
8	0.89	0.89	0.89	9
9	1.00	0.91	0.95	11
11	0.93	0.93	0.93	14
12	1.00	0.86	0.92	14
13	1.00	0.83	0.91	6
15	1.00	1.00	1.00	11
16	0.83	1.00	0.91	10
17	1.00	0.60	0.75	5
18	1.00	1.00	1.00	10
20	0.89	0.89	0.89	9
22	1.00	1.00	1.00	7
23	1.00	1.00	1.00	8
24	1.00	1.00	1.00	8
25	1.00	1.00	1.00	12
26	1.00	1.00	1.00	8
27	0.93	0.93	0.93	14
28	0.90	1.00	0.95	9
32	0.88	1.00	0.94	15
33	0.92	0.86	0.89	14
34	0.90	0.90	0.90	10
35	1.00	1.00	1.00	14
37	1.00	1.00	1.00	7
38	1.00	1.00	1.00	12
39	0.88	1.00	0.93	7
accuracy			0.95	300
macro avg	0.95	0.95	0.95	300
weighted avg	0.95	0.95	0.95	300

Appendix E Classification Report for Cascade Correlation