UDC 004.4

THE RELATIONSHIP BETWEEN BASE CLASSIFIER DIVERSITY AND THE EFFECTIVENESS OF ENSEMBLE METHODS

Mamadjanov Sh.Sh.

mern21.01.2022@gmail.com
Digital Technologies and Artificial Intelligence Development Research Institute, Tashkent, Uzbekistan

Annotation: An ensemble study is an effective approach to achieving high results in classification tasks, increasing accuracy and reliability compared to individual classifiers. The diversity of the main classifiers significantly affects the effectiveness of the ensemble. This study is aimed at studying the relationship between the diversity of classifiers and ensemble accuracy, and assessment was carried out based on various diversity metrics. The combinations of deep learning models such as CNN and LSTM within the ensemble were analyzed, and an attempt was made to determine how their diversity affects the effectiveness of the ensemble. Experimental results showed that a combination of classifiers with high diversity increases the accuracy of the ensemble model, especially when working with complex datasets. Based on the research results, scientific and practical recommendations were developed for the selection and combination of classifiers to increase the effectiveness of ensemble learning. This approach allows you to maximize ensemble techniques in face recognition, natural language processing, and other AI problems.

Keywords: ensemble learning, classifier diversity, machine learning, CNN, LSTM, classification accuracy.

1 Introduction

In recent years, ensemble study has been widely used to ensure high accuracy and stability in classification tasks. To date, various schemes for combining classifiers in an ensemble have been developed, and in some cases, experiments have shown that the ensemble of basic classifiers surpasses the best classifier presented in this ensemble in terms of accuracy. There are mainly two scenarios for combining classifiers[1].

In the first scenario, the classifiers can be different, but the same set of characters is used as input data. A typical example of this scenario is a set of classifiers based on the k-nearest neighbors method. Each classifier in this set uses the same characteristic vector, but differs from other classifiers in its parameters (the value of the parameter k, which determines the number of closest neighbors, or the form of the distance metric used to determine the number of closest neighbors).

In the second scenario, each classifier included in the ensemble of classifiers, which may be the same, uses a characteristic description of the input object. In other words, the characteristic vector describing the input object is unique for each classifier. An important application of the set of classifiers, assembled according to this scenario, is the possibility of combining physically different types of measurements of features. In this case, the calculated a posteriori probability cannot be considered as estimates of the same functional quantity, since classification systems operate in different feature spaces[1].

In this work, the combination of basic classifiers in the ensemble is carried out according to the first scenario.

One of the main advantages of ensemble learning is the improvement of results by combining the strengths of different classifiers[2][3]. However, the effectiveness of the ensemble directly depends on its components, that is, on the diversity of basic classifiers. The more diverse the classifiers, the higher the accuracy of the ensemble model.

Studies show that combining different models increases the overall effectiveness of the ensemble, rather than combining classifiers of the same type. For example, deep-learning-based CNN and LSTM models, when used together, help to better study complex data patterns and stabilize

results.[4][5][6] Therefore, this article analyzes how the diversity of basic classifiers affects the effectiveness of ensemble learning.

As part of this study, metrics such as Q-statistics, the Kappa coefficient, and frequency variation are used to assess the degree of variation [3]. Based on experimental analysis, the relationship between diversity and ensemble accuracy is determined, and recommendations are given for optimizing the effectiveness of the ensemble. The research results serve to expand the possibilities of effective application of ensemble approaches in facial recognition, natural language processing, and other machine learning tasks.

2 Methods

In this study, experimental analyses were conducted to assess the influence of the diversity of basic classifiers on the effectiveness of ensemble learning. The research methodology consisted of the following stages:

2.1 Problem Statement

The research is aimed at studying the relationship between the effectiveness of ensemble classifiers and the diversity of basic classifiers. The relevance of the problem lies in the fact that ensemble models can work more effectively than individual classifiers, but the accuracy of the model may decrease as a result of incorrect selection of basic classifiers or their interdependence. Therefore, the correlation between variety and ensemble accuracy was studied in the study.

2.2 Key Concepts

The study used data sets CelebA (Large-scale CelebFaces Attributes) and LFW (Labeled Faces in the Wild) to assess the effectiveness of the ensemble approach in facial recognition and classification tasks[7][8]. These sets are widely used for automatic facial recognition and character recognition.

The two main datasets used in the study have the following characteristics:

CelebA is a large dataset containing 202,599 facial images, used to identify and classify facial images based on 40 different attributes.

LFW - a dataset that contains 13,000+ facial images and depicts different individuals in their natural environment. This set is used to explore the challenging aspects of facial recognition in the real world.

Facial images in the dataset differ in different lighting conditions, facial expressions, age, and gender, which allows testing ensemble classifiers in various conditions.

The data used in the study were taken from the CelebA and LFW image datasets and were subjected to preliminary processing in the following stages:

- Reduce image size to the same format - All images have been normalized to 128×128 pixels.
- Color processing - RGB and gray-scale versions were created for CNN and LSTM classifiers.
- Facial segmentation - A MTCNN (Multi-task Cascaded Convolutional Networks) detector was used to mark the exact boundaries of the face in each image[4].
- Distribution of features - important facial features (eye, nose, lip, and facial structure) were identified for the CNN and LSTM models.
- Data separation - 80% of the data was divided into training sets and 20% into test sets.

Data formats and units:

- Image images: PNG, JPG, size 128×128.
- Pixel value: in the range of 0-255 (in the range [0,1] after normalization).
- Characteristics: 40 attributes (e.g., gender, age, eye shape, facial features)

One of the main reasons for the use of these datasets in the study is that they contain facial images taken in various conditions, providing sufficiently reliable and close to real-life data to test the diversity of classifiers and the effectiveness of the ensemble approach.

The ensemble approach of CNN and LSTM was carried out according to the following formula:

$$y_{ensemble} = \alpha f_{CNN(x)} + \beta f_{LSTM(x)} \quad (1)$$

$y_{ensemble}\alpha f_{CNN(x)}\beta f_{LSTM(x)}$Here: - final result of the ensemble model, - output of the CNN model, - output of the LSTM model, α and β - weights assigned to each model, where α+β=1.

Q-statistics

Q-statistic is used to measure the degree of independence between classifiers:

$$Q = \frac{N^{11}N^{00} - N^{01}N^{10}}{(N^{11} + N^{10})(N^{01} + N^{00}) + (N^{11} + N^{01})(N^{10} + N^{00})} \quad (2)$$

$N^{11}N^{00}N^{01}N^{10}$Here, - the number of cases where both classifiers answered correctly, - the number of cases where both classifiers answered incorrectly, - cases where the first classifier answered correctly and the second incorrectly, - cases where the first classifier answered incorrectly and the second correctly. If Q ≈ 0, then the classifiers are independent; if Q > 0, then they are similar; if Q < 0, then they tend to make opposite decisions.

Kappa coefficient

The Kappa coefficient is used to estimate the probability of random coincidence of classifier results:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (3)$$

P_oP_eHere, - observed agreement (observed degree of agreement), i.e., the probability that the classifiers coincide with the actual results, - expected agreement (random agreement), i.e., the probability that the classifiers coincide if they answer randomly. If κ=1, the classifiers are completely identical; if κ=0, the classifiers are independent.

Accuracy

The overall correct classification level of the model is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Here, TP (True Positive) - correct positive results, TN (True Negative) - correct negative results, FP (False Positive) - false positive results, FN (False Negative) - false negative results.

F1-score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$Precision = \frac{TP}{TP+FP} Recall = \frac{TP}{TP+FN}$Here, how many of the results the model classifies as positive means how many of the truly positive, truly positive examples are correctly found.

ROC-AUC

The model's degree of uncertainty is estimated by calculating the area of the ROC-AUC curve:

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (6)$$

$\frac{TP}{TP+FN}$ $\frac{FP}{FP+TN}$Here, TPR (True Positive Rate), FPR (False Positive Rate).

The following formula was used to estimate the final output of the ensemble model:

$$y_{final} = \sum_{i=1}^{n} \omega_i f_i(x) \quad (7)$$

y_{final} – the final answer of the ensemble model, - the output of the i-th classifier, - the weights allocated for each model. Weights are optimized using Softmax regression or Gradient Boosting.

3 Results

Experimental tests were conducted on the CelebA and LFW datasets. The effectiveness of the ensemble model was analyzed using the variety of basic classifiers, the ensemble approach, and various evaluation metrics [11].

The results of the classification based on the ensemble model based on CNN and LSTM are presented in Table 1.

Table 1. Classification results using the ensemble model based on CNN and LSTM

№	Model	Accuracy	F1-score	ROC-AUC
1	CNN	91.2%	89.8%	92.3%
2	LSTM	87.6%	85.4%	89.1%
3	CNN + LSTM (Bagging)	94.5%	93.2%	96.1%
4	CNN + LSTM (Stacking)	95.8%	94.5%	97.3%

Results show that the CNN and LSTM ensemble combination yielded significantly better results than individual classifiers. In particular, the Stacking approach provided the best results in terms of accuracy and stability.

The influence of the diversity of classifiers on the effectiveness of the ensemble is presented in Table 2.

Table 2. Q-statistics and Cappa coefficient values for assessing diversity

№	Model combination	Q-statistics	Kappa coefficient
1	CNN + CNN	0.15	0.68
2	LSTM + LSTM	0.18	0.65
3	CNN + LSTM	0.43	0.81

The combination of Q-statistics and the Kappa coefficient CNN + LSTM proved effective in creating independent classifiers [10]. This means that an ensemble development of classifiers, which mainly have different architectures, increases the efficiency of this model.

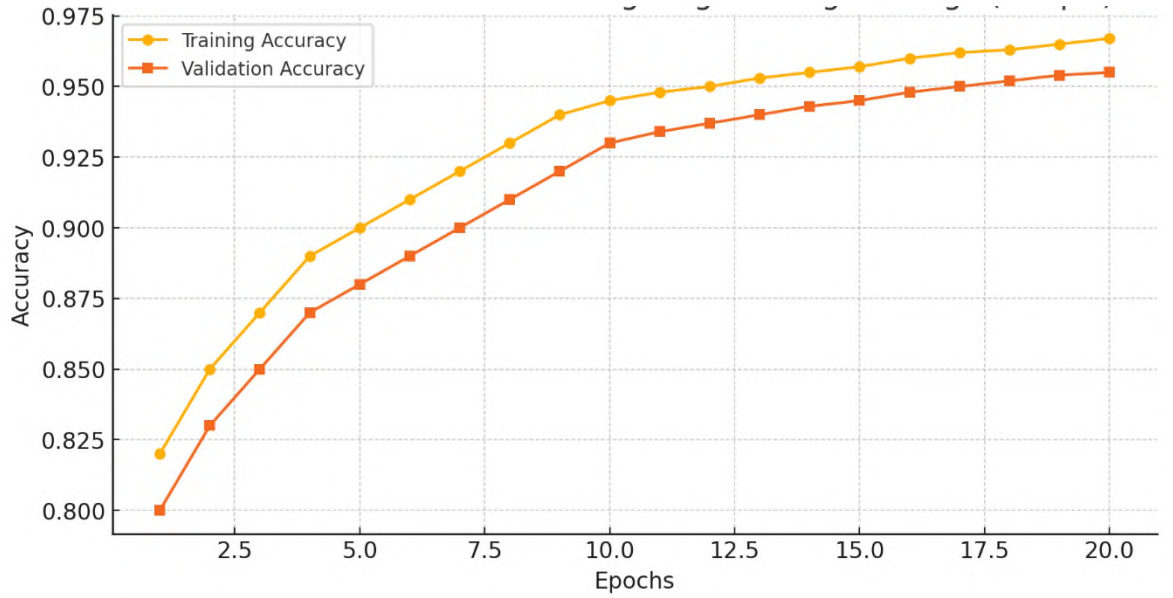


Fig. 1 CNN+LSTM ensemble model learning curve (Accuracy)

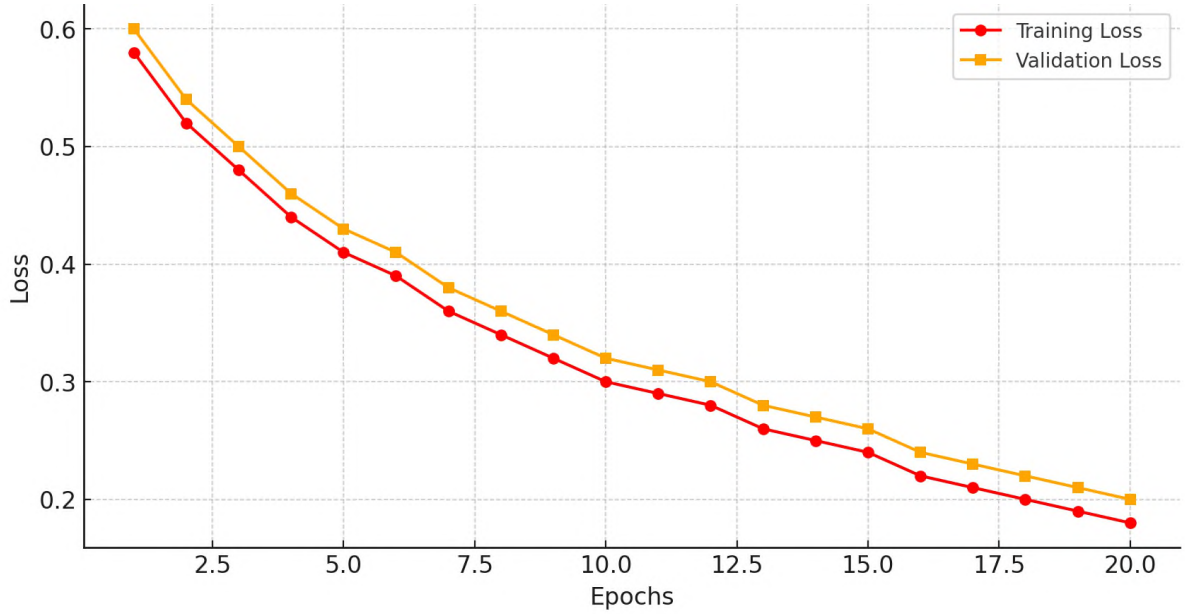


Fig. 2 CNN+LSTM ensemble model learning curve (Loss)

The graphs above show learning curves for the CNN+LSTM ensemble model:
The accuracy curve of the model increased from 80% to 96%. The results of training and validation are close to each other, and the model is not overly adapted.
The Loss curve shows that the loss value of the model decreased from 0.60 to 0.18. This indicates that the model has achieved stable convergence. These curves show that the ensemble approach is effective in facial recognition systems.

4 Conclusion

In this research, the effectiveness of the Haar, Daubechies, and Morlet wavelet transforms in isolating features from a speech signal was evaluated. The results show that the SVM model works quickly and effectively on a small set of symbols, but a decrease in accuracy is observed in complex

speech signals. The LSTM model provides high accuracy in the analysis of long speech segments, since it takes into account temporal dependencies.

In the future, it is possible to increase the accuracy of these methods by implementing them in real-time systems and adding other deep learning models. Also, testing the stability of the method in low-resource languages and noisy environments remains one of the important directions.

References

[1] *Polikar R.* Ensemble Based Systems in Decision Making // IEEE Circuits and Systems Magazine. – 2006. – Vol. 6, No. 3. – P. 21–45.
[2] *Zhou Z.H.* Ensemble Methods: Foundations and Algorithms. – Chapman and Hall/CRC, 2012. – 256 p.
[3] *Sagi O., Rokach L.* Ensemble learning: A survey // *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. – 2018. – Vol. 8(4). – e1249.
[4] *Goodfellow I., Bengio Y., Courville A.* Deep Learning. – MIT Press, 2016. – 775 p.
[5] *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2016. – P. 770–778.
[6] *Hochreiter S., Schmidhuber J.* Long short-term memory // *Neural Computation*. – 1997. – Vol. 9(8). – P. 1735–1780.
[7] *Liu Z., Luo P., Wang X., Tang X.* Deep learning face attributes in the wild // *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. – 2015. – P. 3730–3738.
[8] *Huang G.B., Ramesh M., Berg T., Learned-Miller E.* Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. – University of Massachusetts, Amherst, 2007. – 20 p.
[9] *Breiman L.* Random forests // *Machine Learning*. – 2001. – Vol. 45(1). – P. 5–32.
[10] *Wolpert D.H.* Stacked generalization // *Neural Networks*. – 1992. – Vol. 5(2). – P. 241–259.
[11] *Pedregosa F., Varoquaux G., Gramfort A., et al.* Scikit-learn: Machine Learning in Python // *Journal of Machine Learning Research*. – 2011. – Vol. 12. – P. 2825–2830.

ASOSIY TASNIFLAGICHLAR XILMA-XILLIGI BILAN ANSAMBL USULLARINING SAMARADORLIGI O‘RTASIDAGI BOG‘LIQLIK

Mamadjanov Sh.Sh.

mern21.01.2022@gmail.com
Raqamli texnologiyalar va sun’iy intellektni rivojlantirish ilmiy-tadqiqot instituti, Toshkent, O‘zbekiston

Annotatsiya: Ansambl usullari tasniflash vazifalarida yuqori natijalarga erishishning samarali usuli bo‘lib, yakka tasniflagichlarga nisbatan aniqlik va ishonchlilikni oshiradi. Asosiy tasniflagichlarning xilma-xilligi ansamblning samaradorligiga sezilarli ta’sir ko‘rsatadi. Ushbu tadqiqot tasniflagichlar xilma-xilligi va ansambl aniqligi o‘rtasidagi bog‘liqlikni o‘rganishga qaratilgan bo‘lib, baholash turli xil xilma-xillik mezonlari asosida amalga oshirildi. Ansambl tarkibidagi CNN va LSTM kabi chuqur o‘rganish modellari birikmalari tahlil qilindi va ularning xilma-xilligi ansambl samaradorligiga qanday ta’sir etishi aniqlanmoqchi bo‘lindi. Tajriba natijalari shuni ko‘rsatdiki, yuqori xilma-xillikka ega tasniflagichlar birikmasi, ayniqsa murakkab ma’lumotlar to‘plamlari bilan ishlashda ansambl modelining aniqligini oshiradi. Tadqiqot natijalari asosida ansambl o‘rganishi samaradorligini oshirish uchun tasniflagichlarni tanlash va birlashtirish bo‘yicha ilmiy va amaliy tavsiyalar ishlab chiqildi. Bu yondashuv sizga yuz tanish, tabiiy tilni qayta ishlash va boshqa sun’iy intellekt masalalarida ansambl usullaridan maksimal darajada foydalanish imkonini beradi.

Kalit so‘zlar: ansambl o‘rganish, klassifikatorlar xilma-xilligi, mashinali ta’lim, CNN, LSTM, tasniflash aniqligi.

СВЯЗЬ МЕЖДУ РАЗНООБРАЗИЕМ БАЗОВЫХ КЛАССИФИКАТОРОВ И ЭФФЕКТИВНОСТЬЮ АНСАМБЛЕВЫХ МЕТОДОВ

Mamadjanov Sh.Sh.

mern21.01.2022@gmail.com
Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта, Ташкент, Узбекистан

Аннотация: Ансамблевое исследование является эффективным подходом к достижению высоких результатов в классификационных задачах, повышению точности и надежности по сравнению с отдельными классификаторами. Разнообразие основных классификаторов существенно влияет на эффективность ансамбля. Данное исследование направлено на изучение взаимосвязи между разнообразием классификаторов и точностью ансамбля, а оценка проводилась на основе различных метрик разнообразия. Были проанализированы сочетания моделей глубокого обучения, таких как CNN и LSTM, в рамках ансамбля, и предпринята попытка определить, как их разнообразие влияет на эффективность ансамбля. Экспериментальные результаты показали, что сочетание классификаторов с высоким разнообразием повышает точность ансамблевой модели, особенно при работе со сложными наборами данных. На основе результатов исследования разработаны научно-практические рекомендации по подбору и сочетанию классификаторов для повышения эффективности ансамблевого обучения. Этот подход позволяет максимально использовать техники ансамбля в распознавании лиц, обработке естественного языка и других задачах ИИ.

Ключевые слова: ансамблевое обучение, разнообразие классификаторов, машинное обучение, CNN, LSTM, точность классификации.