



Role of Chi-Square in Enhancing the Accuracy of Classifiers in Emotion Recognition

Pushpa Pathak^{1*}

¹FCA, Acropolis Institute of Technology and Research, RGPV, Indore, Madhya Pradesh, India

*pushpapathak@acropolis.in

Abstract. Emotion recognition has been a fast-emerging field in artificial intelligence. It has applications in the healthcare sector, human-computer interaction, and behavioral studies. Feature selection is vital in improving the accuracy and efficiency of classifiers in this field, especially for high-dimensional data. This paper investigates the utilization of the chi-square test for improving the performance of the various classifiers for emotion recognition. The experiment uses the benchmarking CK+ (Cohn-Kanade) database that contains labeled facial expression images. The chi-square test is used to select features in the most relevant manner corresponding to Action Units (AU), related to emotions, targeting here "sadness". Reduced dimensionality of the dataset gives rise to the chi-square test while removing irrelevant features and alleviating noise and computational complexity associated with it. Accuracy obtained from the K-Nearest Neighbour (k-NN) classifier is compared before as well as after applying chi-square feature selection. The preliminary results show high improvement in the classification accuracy score, 79% becoming 87%, which indicates high effectiveness in refining the features by the application of chi-square. Thus, this experiment shows how statistical feature selection techniques have relevance in optimizing various machine learning classifiers for accurate emotion recognition. After observing the improvement in the performance of KNN, the performance of classifiers like Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR) and Naïve Bayes (NB) was also checked. Improvement in F1-score for each classifier was recorded after applying chi-square statistical tool on the relevant extracted features from the images. The possible combination of the chi-square attribute with classifiers can really lead to robust and highly efficient systems in emotion recognition technology, leading to more high-end applications of AI-powered systems.

Keywords Emotion Recognition, Action Units, Chi-Square, K-Nearest Neighbour, Random Forest, Support Vector Machine, Logistic Regression, Naïve Bayes

1 Introduction

Emotion recognition is an emerging field that enhances the interaction between humans and machines. Accurately identifying emotions such as sadness plays a pivotal role in applications ranging from mental health monitoring to adaptive systems. However, the high dimensionality of facial features[1], particularly Action Units (AUs) [2], often hampers classifier efficiency and accuracy. Feature selection techniques like the chi-square test offer a solution by reducing the dataset's dimensionality while retaining relevant features. This study investigates the role of chi-square in optimizing the performance of classifiers[3] for sadness recognition using the CK+ dataset [4], emphasizing the importance of prioritizing features for effective classification. The chi-square (χ^2) test is a statistical method used to examine relationships between categorical variables [5]. In classification tasks, particularly those involving categorical data, the chi-square test plays an essential role in feature selection. Chi-square can help determine which features (predictors) are more relevant to the target variable. In classification problems, it's important to pick features that have a strong relationship with the output class. The chi-square test assesses whether a feature's distribution of categories is independent of the target variable's categories. Features with a lower p-value (below a significance threshold) are considered to have a strong relationship with the class label, and are more likely to contribute to accurate classification.

$$\chi^2 = \sum (O_i - E_i)^2 / E_i \quad (1)$$

where

O_i: Observed frequency of a category.

E_i: Expected frequency of a category .

In facial expression analysis, "Action Units" (AUs) are used to describe the movements of different facial muscles. These movements correspond to changes in facial appearance, and specific combinations of AUs are associated with various emotions, such as sadness, happiness, anger, etc.[6]. The Facial Action Coding System (FACS), developed by Paul Ekman and Wallace Friesen, provides a standardized way to label these AUs. For our study we have considered the sad emotion and the classifier used is KNN. According to

Ekman and Friesen for sad emotion the following action units play a vital role in identifying the emotion label[2].

Primary Action Units for Sadness:

AU 1: Inner Brow Raiser

Involves lifting the inner part of the eyebrows, creating a furrowed appearance.

Often seen as the most prominent indicator of sadness.

AU 4: Brow Lowerer

Lowers the eyebrows, creating a frown-like appearance.

Enhances the expression of distress and sadness.

AU 15: Lip Corner Depressor (or Lip Corner Puller)

Pulls down the corners of the lips, forming a slight frown.

This AU is particularly important for the typical "sad" mouth shape.

Secondary Action Units for Sadness:

AU 11: Nasolabial Deepener

Pulls the corners of the nose downwards, making the nasolabial folds (lines from the nose to the corners of the mouth) more pronounced.

Can enhance the intensity of the sad expression.

AU 6: Cheek Raiser

Slightly raises the cheeks, often seen in conjunction with tears or sadness when combined with squinting.

AU 54: Head Down (Head Movement)

Tilting the head downwards can accompany a sad expression, adding to the overall impression of a downcast, subdued demeanor.

AU 1 + AU 4 + AU 15: This combination of raised inner brows, lowered brows, and depressed lip corners is often sufficient to communicate sadness.

AU 1 + AU 4 + AU 15 + AU 11: Adds more detail to the sad expression, especially if the sadness is intense.

2 Literature Review

Ekman and Friesen's introduced Action Units (AUs) as a standardized way to measure facial expressions. Several emotion recognition systems leverage these AUs for feature

extraction. K-NN classifier is widely used in emotion recognition due to its simplicity and effectiveness in non-linear problems. However, its performance is sensitive to irrelevant or redundant features. Techniques like Principal Component Analysis (PCA), Ada Boost and chi-square have been employed to enhance classifier performance [7]. Ada Boost extracts the features, PCA reduces dimensionality by transforming features, chi-square ranks features based on statistical significance, making it particularly useful for categorical data.

Sadegh-Zadeh et al. [8] have concluded that the utilization of dimensionality reduction techniques can result in better classification of EEG data since they reduce computational demands and achieve it without loss of accuracy. The results open the possibility of using such techniques in scenarios where high accuracy and increased efficiency are desired.

Hokijuliandy E. et. al.[9] in their study present an effective SVM classification approach as well as Chi-Square feature selection in the process of sentiment analysis. Findings were valuable for improving the understanding of users' sentiment regarding the Indonesia's National Health Insurance (Mobile JKN) application, thus leading to further user experience improvement and progress in sentiment analysis.

Dong Liu [10] with other authors, proposed that the method first adopts the chi-square test method to delete the redundant information and noise information, then apply emotion recognition processing to the features selected from the mono-modal feature. The two modalities then integrated to recognize the emotions. A set of experiments has been designed to compare the model performance with and without feature selection based on the chi-square test.

Chuanyu Tang in his paper [11] outlines some applications of the chi-square statistic applied to text classification from the last five years and focuses on applying it to Arabic text, social media data, and medical research. The chi-square statistic has presented prominent merits in text classification tasks and has enhanced the performance of classifiers in linguistically complex Arabic texts, a mass amount of user-generated content within social media data, and the classification of medical literature. In the near future, researchers may employ other statistics besides the chi-square statistic to enhance the precision of the classification of texts according to the changing needs of this field.

Y. Zhai et. al. [12] have proposed a method for extracting feature words based on Chi-square Statistics. Feature words are co-occurring or appear independently depending on different circumstances. The authors in their research have considered the text using both single word and double words as features. Experimental results were compared and

analyzed in detail to illustrate the effectiveness of their method using Naive Bayes and Support Vector Machine classifiers.

3 Methodology

3.1 Dataset

The CK+ dataset holds approximately 8,000 to 10,000 images in total, but the number of images used depends on the specific experimental protocol. In the case of sadness detection, the number of frames selected per sequence may influence the final size of the dataset. The CK+ dataset, consisting of 593 labeled sequences[4] of facial expressions, is used in the experiment. Each sequence represents a number of frames(images) taken over time as the person's expression transits from neutral expression to peak expression. Usually each sequence consists of 10 to 60 frames. The number exactly depends upon recording conditions and expression progression. Thus on an average 15 frames per sequence are taken. Therefore in the study we have $593 * 15 = 8895$ images. Generally peak frames with maximum expression are taken. The dataset includes AU annotations and emotion labels, with sadness being the target emotion in this study.

3.2 Face Detection

Viola-Jones face detection technique is used to extract facial regions from images.

3.3 Feature Extraction

Action Units (AUs) are extracted using a pre-trained AdaBoost-based model[7] as shown in table 1. These AUs quantify specific facial muscle movements. The extracted features are shown in the table. Here every row represents the intensity of every labeled action unit for a facial image. The data set contains a large number of AUs some of which may be irrelevant or redundant for emotion recognition.

Table 1. All the extracted action units(AUs)

| AU1 | AU2 | AU3 | AU4 | AU5 | | AU15 | AU16 |
|-----|-----|-----|-----|-----|-------|------|------|
| 0.8 | 0.4 | 0.7 | 0.9 | 0.2 | | 0.6 | 0.3 |
| 0.3 | 0.5 | 0.2 | 0.4 | 0.8 | | 0.1 | 0.6 |

0.9 0.3 0.5 0.7 0.1 0.4 0.2

But when Chi-Square is applied to the selected AUs then, only the most prominent AUs are prioritized and given to the classifier for emotion classification. The chi-square test is applied to rank AUs based on their statistical significance in identifying sadness. Irrelevant AUs are discarded, retaining a prioritized set of features most relevant to sadness.

In the process after the action units are extracted by feature extractor Ada Boost, chi-square test is applied to a dataset where each row is a facial image. After applying the chi-square test, action units are prioritized in the following order

Table 2. Prioritized action unit using Chi-Square

| Action Unit | Chi-Square Score | p-value | Priority order |
|-------------|------------------|---------|----------------|
| AU1 | 34.5 | <0.001 | 1 |
| AU4 | 28.7 | <0.001 | 2 |
| AU15 | 22.3 | <0.01 | 3 |
| AU11 | 15.4 | <0.05 | 4 |

As shown in table 2 AU1 is given the highest priority as it has the highest chi-square score and lowest p-value, indicating it is the most relevant feature for the sadness label. Similarly the prioritization of action units AU4, AU15 and AU11 is done.

3.4 Classification

In this step the dataset is given to classifier K-NN to identify the emotion label. Therefore the classifier checks the intensity of all the extracted action units and then identifies the label with 1 or 0 as shown in the table 3 below.

Table 3. Emotion identification with all the available action units

| AU1 | AU2 | AU3 | AU4 | AU5 | | AU15 | AU16 | Emotion Label |
|-----|-----|-----|-----|-----|-------|------|------|---------------|
| 0.8 | 0.4 | 0.7 | 0.9 | 0.2 | | 0.6 | 0.3 | 1 |
| 0.3 | 0.5 | 0.2 | 0.4 | 0.8 | | 0.1 | 0.6 | 0 |
| 0.9 | 0.3 | 0.5 | 0.7 | 0.1 | | 0.4 | 0.2 | 1 |
| 0.0 | 0.2 | 0.6 | 0.2 | 0.0 | | 0.1 | 0.8 | 0 |

But with chi-square implementation only the most important action units and their intensities are considered in identifying the label. This was possible because chi-square prioritizes the action units. For example, here in table 4, for sad emotion the intensity of action units AU1, AU4, and AU15, are only considered.

Table 4. Emotion Identification after prioritization action units by Chi-Square

| AU1 (Inner Brow Raiser) | AU4 (Brow Raiser) | AU15 (Lip Corner Depressor) | Emotion Label |
|----------------------------|----------------------|-----------------------------------|------------------|
| 0.8 | 0.9 | 0.6 | 1 |
| 0.3 | 0.4 | 0.1 | 0 |
| 0.9 | 0.7 | 0.4 | 1 |
| 0.0 | 0.2 | 0.0 | 0 |

Each row represents a sample (e.g., a facial image), and the columns are AU values representing feature intensity. The emotion label 1, indicates that the emotion depicted in the sample image is sad. Likewise the emotion value 0 represents not sad.

The extracted features were also given to Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR) and Naïve Bayes (NB) classifiers before and after Chi-Square prioritization of action units.

4 Performance Evaluation

Metrics such as accuracy, precision, recall, and F1-score are computed to evaluate k-NN's performance before and after chi-square feature selection.

The results of the k-NN classifier can be represented in a confusion matrix (Table 5)

Table 5. Confusion matrix for sad emotion prediction

| Emotion | Predicted sad | Predicted not sad |
|---------|---------------|-------------------|
| Sad | 50 (TP) | 10 (FN) |
| Not Sad | 5(FP) | 35 (TN) |

TP = 50 (Correctly predicted sadness)

TN = 35 (Correctly predicted not sadness)

FP = 5 (Incorrectly predicted sadness)

FN = 10 (Incorrectly predicted not sadness)

Given the values of TP, TN, FP, and FN, Accuracy, Precision, Recall, F1 Score, False positive rate and specificity can be calculated as:

$$\text{Accuracy} = (50+35)/(50+35+5+10)=85/100=0.85 \text{ or } 85\%$$

$$\text{Precision} = 50/(50+5) = 50/55 = 0.91 \text{ or } 91\%$$

$$\text{Recall} = 50/ (50+10)=50/60=0.83 \text{ or } 83\%$$

$$\text{F1-Score} = (2 \times 0.91 \times 0.830) / (0.91+0.83)=1.51/1.74 =0.87 \text{ or } 87\%$$

$$\text{False Positive Rate (FPR)} = 5/(5+35) =5/40 =0.125 \text{ or } 12.5\%$$

$$\text{Specificity}=3/(55+35) = 35/40=0.875 \text{ or } 87.5\%$$

91% is high precision which indicates that when the model predicts "sadness," it is correct 91% of the time. Moderate Recall 83% is moderate recall which indicates that the model can detect 83% of all actual "sadness" cases. 87% F1-Score is a balance between precision and recall, suggesting good performance in terms of both false positives and

false negatives. These metrics (Table 6) collectively provide a detailed picture of how well the k-NN classifier is performing in identifying sadness.

Table 6. Performance matrix of k-NN classifier

| Metric | Before Chi-Square | After Chi-Square |
|-----------|-------------------|------------------|
| | (All AUs) | (Top 4 AUs) |
| Accuracy | 79% | 87% |
| Precision | 68% | 76% |
| Recall | 60% | 67% |
| F1-Score | 64% | 71% |

Following the same steps as above, we in our experiment also considered RF, SVM, LR and NB and found the following results in terms of accuracy and F1- score for the classifiers are shown below in table 7.

Table 7. Performance matrix for all the classifiers

| Classifier | Accuracy | Accuracy | F1-Score | F1-Score |
|------------|----------|----------|----------|----------|
| | (Before) | (After) | (Before) | (After) |
| KNN | 79% | 87% | 64% | 71% |
| RF | 81% | 88% | 66% | 73% |
| SVM | 77% | 85% | 62% | 70% |
| LR | 75% | 83% | 61% | 68% |
| NB | 73% | 81% | 59% | 67% |

5 Discussion

The results demonstrate the efficacy of the chi-square test in optimizing the efficiency of classifiers in recognizing the sad emotion. By prioritizing relevant AUs, the chi-square test not only improves accuracy but also enhances computational efficiency. The findings underscore the importance of feature selection in addressing the challenges posed by high-dimensional datasets. When the chi-square test is applied to rank Action Units (AUs) based on their importance in emotion recognition, it makes the classification process significantly more efficient and accurate for any classifier. The increase in F1-score for all the classifiers, indicate that the irrelevant or noisy features were removed, allowing the model to focus on the most informative features. An accuracy of 87% means that, out of 100 predictions made by the k-NN classifier, approximately 87 predictions were correct, and 13 predictions were incorrect. Before chi-square implementation, all the AUs were checked. The model struggled with unnecessary complexity and noise, which made it harder to correctly identify sadness. But after Chi-Square prioritization of action units, with fewer, more relevant AUs, classifiers worked more effectively, improving accuracy, precision, and recall with different degree of improvement. k-NN and Logistic Regression are the most improved because they are very sensitive to feature selection. Random Forest and Naive Bayes also improve but are inherently more robust, so the effect of chi-square is less significant. SVM benefits from clear boundaries in feature space after elimination of irrelevant action units.

6 Conclusion

This study highlights the critical role of feature selection in emotion recognition. By applying the chi-square test to prioritize AUs, the classifiers achieved improved accuracy and efficiency in recognizing sadness using the CK+ dataset. The results suggest that statistical feature selection techniques can significantly enhance the performance of machine learning models in affective computing. Emotion recognition is a critical aspect of affective computing with applications in healthcare, human-computer interaction, and behavioral studies. This research focuses on optimizing the process of identifying the emotion of sadness using the various classifiers. The optimization is achieved by employing the chi-square test to prioritize Action Units (AUs) critical for recognizing sadness. The CK+ (Cohn-Kanade) dataset is utilized for experimentation, where extracted features are ranked using the chi-square test and fed into the classifiers. Results indicate a significant improvement in classification accuracy and computational efficiency,

demonstrating the potential of combining chi-square feature selection with classifiers for robust emotion recognition systems.

Chi-Square ranks AUs based on their statistical significance for classifying emotions like sadness. The priority order of AUs is based on the chi-square scores, with higher scores indicating stronger associations between the AU and the emotion (e.g., sadness).

AUs with the highest chi-square scores are selected as the key features for classifying the emotion, improving the model's focus on the most relevant facial features.

It was found that the best impact of applying chi-square is on simpler models like k-NN and LR experience a greater impact since these algorithms are more susceptible to irrelevant features and noise. Chi-square enables them to consider only the most relevant features, thereby improving decision-making. Classifiers like RF and NB are already relatively robust to irrelevant features. While they do improve post chi-square, the impact is less dramatic compared to simpler models. RF naturally handles many features via feature bagging, and NB assumes feature independence, so the chi-square test plays a smaller role in improving these classifiers. SVM shows noticeable improvement, especially in separating emotion classes, as the clearer margin between classes post-feature selection allows for better decision boundary formation.

Future research can expand this approach to other datasets and emotions, paving the way for more robust emotion recognition systems.

References

1. Viola, P. and Jones, M., Rapid Object Detection Using a Boosted Cascade of Simple Features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, 8-14 December 2001, 1-511-1-518. [https://doi.org/10.1109/CVPR.2001.990517_\(2001\)](https://doi.org/10.1109/CVPR.2001.990517_(2001))
2. Ekman, P., Friesen, W., "Facial Action Coding System: A Technique for the Measurement of Facial Movements", Consulting Psychologists Press, California, 1978.
3. Cover T., Hart, P.,: "Nearest neighbor pattern classification," in IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, January 1967, doi: 10.1109/TIT.1967.1053964.
4. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I.,: "The extended cohn-kanade dataset (CK) : A complete dataset for action unit and emotion-specified expression," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, 2010, pp. 94–101.

5. Onchiri, S. Conceptual model on application of chi-square test in education and social sciences. *Education research and reviews*, 8(15), 1231-1241.(2013)
6. Pushpa T. and Veenadhari S., "An Efficient classification technique for automatic identification of emotions leading to stress" IEEE 6th conference on Information and communication Technology(CICT-2022)
7. Bishop, C.M. *Pattern Recognition and Machine Learning*. Springer, Berlin. [https://link.springer.com/book/9780387310732_\(2006\)](https://link.springer.com/book/9780387310732_(2006))
8. Sadegh-Zadeh SA, Sadeghzadeh N, Soleimani O, Shiry Ghidary S, Movahedi S, Mousavi SY.; Comparative analysis of dimensionality reduction techniques for EEG-based emotional state classification. *Am J Neurodegener Dis*. 2024 Oct 25;13(4):23-33. doi: 10.62347/ZWRY8401. PMID: 39584052; PMCID: PMC11578865.
9. Hokijuliandy, E., Napitupulu, H., Firdaniza.; Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia's National Health Insurance Mobile Application. *Mathematics*, 11,3765. [https://doi.org/10.3390/math11173765\(2023\)](https://doi.org/10.3390/math11173765(2023)).
10. Liu D, Wang Z, Wang L and Chen L, Multi-Modal Fusion Emotion Recognition Method of Speech Expression Based on Deep Learning. *Front. Neurobot*. 15:697634. doi: 10.3389/fnbot.2021.697634,(2021).
11. Chuanyu, T., "Review on Application of Chi-square Statistic in Text Classification in Recent Five Years", *Proceedings of the 2nd International Conference on Machine Learning and Automation*. DOI: 10.54254/2755-2721/97/20241397,(2018).
12. Zhai, Y., Song, W., Liu, X., Liu, L., and Zhao, X., : A Chi-Square Statistics Based Feature Selection Method in Text Classification, in *IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, pp. 160-163, doi: 10.1109/ICSESS.2018.8663882,2018
13. Barigou, F., Improving K-nearest neighbor efficiency for text categorization , *Neural Network World*, 26(1), 45,(2016)
14. Chen, Y. W., Zhou, Q., Luo, W., & Du, J. X., : Classification of Chinese texts based on recognition of semantic topics. *Cognitive Computation*, 8(1), 114-124,2016
15. Kilimci, Z. H., S. Akyokus, S., & S. I. Omurca, S. I., : The effectiveness of homogenous ensemble classifiers for Turkish and English texts, In *International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)* (pp. 1-7). IEEE. (2016)
16. Bahassine,, S., Madani, A., Al-Sarem, M., & Kissi, M., : Feature selection using an improved Chi-square for Arabic text classification, *Journal of King Saud University-Computer and Information Sciences*, 32(2), 225-231,(2020)
17. Alshaer, H. N., Otair, M.A., Abualigah, L., Alshinwan, M. & Khasawneh A. M., : Feature selection method using improved CHI Square on Arabic text classifiers: analysis and application, *Multimedia Tools and Applications*, 80, 10373-10390,(2021)
18. Falasari, A., & Muslim, M.A., : Optimize naïve bayes classifier using chi square and term frequency inverse document frequency for amazon review sentiment analysis, *Journal of Soft Computing Exploration*, 3(1), 31-36,(2022)
19. Paudel, S., Prasad, P., Alsadoon, W.C., Islam, A., Elchouemi, M.R, Feature selection approach for Twitter sentiment analysis and text classification based on Chi-Square and Naïve Bayes, In *International Conference on Applications and Techniques in Cyber Security and Intelligence*,(2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

