The Scientific Journal of King Faisal University

Basic and Applied Sciences

Deep Capsule Network for Facial Emotion Recognition

Tegani Salem [1] and Telli Abdelmoutia [2]

[1] LESIA Laboratory, Department of Electrical Engineering, Faculty of Science and Technology, Biskra University, Biskra, Algeria
[2] Department of Computer Science, Faculty of Exact Sciences and Natural and Life Sciences, Biskra University, Biskra, Algeria

LINK	RECEIVED	ACCEPTED	PUBLISHED ONLINE	ASSIGNED TO AN ISSUE
https://doi.org/10.37575/b/cmp/210063	08/09/2021	30/11/2021	30/11/2021	01/12/2021
NO. OF WORDS	**NO. OF PAGES**	**YEAR**	**VOLUME**	**ISSUE**
5799	6	2021	22	2

ABSTRACT

Although the classification of images has become one of the most important challenges, neural networks have had the most success with this task; this has shifted the focus towards architecture-based engineering rather than feature engineering. However, the enormous success of the convolutional neural network (CNN) is still far from comparable to the human brain's performance. In this context, a new and promising algorithm called a capsule net that is based on dynamic routing and activity vectors between capsules appeared as an efficient technique to exceed the limitations of the artificial neural network (ANN), which is considered to be one of the most important existing classifiers. This paper presents a new method-based capsule network with light-gradient-boosting-machine (LightGBM) classifiers for facial emotion recognition. To achieve our aim, there were two steps to our technique. Initially, the capsule networks were merely employed for feature extraction. Then, using the outputs computed from the capsule networks, a LightGBM classifier was utilised to detect seven fundamental facial expressions. Experiments were carried out to evaluate the suggested facial-expression-recognition system's performance. The efficacy of our proposed method, which achieved an accuracy rate of 91%, was proven by its testing the results on the CK+ dataset.

KEYWORDS

Image classifications, LightGBM, machine learning, computer vision, CNN, deep learning

CITATION

Salem, T. and Abdelmoutia, T. (2021). Deep capsule network for facial emotion recognition. *The Scientific Journal of King Faisal University: Basic and Applied Sciences*, **22**(2), 130–5. DOI: 10.37575/b/cmp/210063

1. Introduction

Some tasks that are considered simple for the human brain, such as facial recognition, detection and segmentation, present a challenging problem for computer vision systems. These vision systems are created using predictive classification modelling, and they made progress in terms of the development of robust systems over the last decade with regard to use of this natural type of human communication (Black and Yacoob, 1995; Essa and Pentland, 1997; Terzopoulos and Waters, 1990; Yacoob and Davis, 1996). Nowadays, tasks relating to computer vision require efficiency at solving common problems like facial recognition, detecting objects, translating languages, age estimation and object segmentation. Even classical artificial intelligence and all of its complicated functions and instructions cannot solve these complex problems, which has led to the creation of new models of deep learning (Mellouk and Handouzi, 2020), such as CNNs.

However, CNNs experience considerable difficulties when trying to recognise small datasets, different poses and deformed objects, even though they require a lot of data for training. As a result of these challenges, a new architecture has been invented within the field of deep learning called capsule networks. They have met expectation levels as they have outperformed CNNs in relation to solving the problems mentioned above and giving highly accurate results in various fields (Hong *et al.*, 2021; Tiwari and Jain, 2021).

In this study, the system of inputting images into our model has been implemented, which is a new method based on deep learning for the detection of facial expressions to predict seven main facial expressions: fear, anger, surprise, happiness, sadness, contempt and disgust.

2. Related Works

In recent years, novel recognition frameworks (Kim *et al.*, 2015; Liu *et al.*, 2014; Ranzato *et al.*, 2011) that have depended on the use of a CNN have produced impressive performances in terms of facial-expression-recognition systems; they have also been utilised for object recognition and feature extraction. A CNN with many convolutions and pooling layers can extract multi-level and higher features from the local area or the full face, and they perform well in relation to facial-emotion-picture-feature classification. Furthermore, a range of convolutional neural-network architectures (Ko, 2018) has been modelled in several studies to identify emotions after the successful introduction of CNNs for various computer vision tasks. A CNN can extract features automatically, capturing all potential complicated non-linear relationships between them. They have also been demonstrated to have promising capabilities for emotion categorisation, as shown in certain studies (Mehendale, 2020; Minaee *et al.*, 2021; Valdenegro-Toro *et al.*, 2019).

Some techniques (Minaee *et al.*, 2021; Valdenegro-Toro *et al.*, 2019; Zeng *et al.*, 2018) concentrate on creating new classifiers and feature extractors for emotion classification (Kim *et al.*, 2013). Substituting the softmax layer with a classifier at the last step in the model of deep learning allows for finer tweaking of the lower-level features. However, because a CNN's underlying data representation ignores crucial spatial hierarchies between complex and simple objects, it cannot accomplish rotational invariance. Nonetheless, when it comes to facial emotion identification, a facial picture can be rotated or translated. The presence of a part is noted by the max-pooling layer in a CNN but not the spatial relationship between the parts themselves. As a result, there is no pose connection between lower-level features that make up a higher-level feature. On the other hand, this connection is critical when developing solid high-level features that help categorisation; the goal is to present a network that can simulate an image's hierarchical connections.

With regard to the extraction of features after studying works by Patrick *et al.* (2019), Tereikovska *et al.* (2019) and Zhang and Xiao, (2020), an architecture based on the capsule network is suggested,

Corresponding Author: Tegani Salem 00213663544420, salem.tegani@univ-biskra.dz

which is a collection of neurons' outputs that reflect several features of the same thing. A capsule network has several layers, each of which includes numerous capsules that were first presented by Sabour *et al.* (2017). In the last step of the facial-emotion-recognition (FER) system, the LightGBM classifier is applied to avoid all of the problems associated with CNNs, leading to a robust model for this system.

3. Description and Backgrounds

3.1. Facial Emotion Recognition:

Humans need to convey their intentions and emotional state so that they can interact with their environment, and facial emotions are a natural way and the universal language that humans use for this purpose. Darwin's work was responsible for the first examination of emotion-based signals as expressed by human faces (Darwin and Prodger, 1998). Several studies have been conducted concerning FER because of its utility in many fields and other systems based on human–computer interactions, such as robotics and gaming, marketing, criminal interrogations, biometric technology and surveillance systems. In the last century, Friesen and Ekman (1976), Ekman (1993), Matsumoto (1992) completed research in relation to the FER phenomenon. The majority of the systems they employed attempted to recognise six prototypic emotion categories: happiness/disgust, anger/surprise and sadness/fear. Contempt was added later on in 1986 to the set of basic facial expressions.

There are two main categories for FER systems based on their feature representations (Corneanu *et al.*, 2016): dynamic sequences and static images (spatial information used to represent features). Starting with these two methods, multimodal systems have used many audio approaches as well as electrocardiograms (ECGs) and electroencephalographs (EEGs) to assist with the recognition of emotion. Leading on from this, nonverbal communication involves facial expressions. Factors such as tone of voice and the context of the words in the argument may distract the investigator and divert their attention away from observing the subject's facially expressed emotion. However, the technology involved in automatic facial emotion recognition systems is not impacted by contextual interference. Medical treatment systems, psychiatric care, driver-fatigue detectors, and computer-animation technology have made gains from the implementation of automatic FER methods. The seven basic emotions categories are attempted to be identified by the technology involved in facial emotion recognition systems.

FER is based on three techniques (Corneanu *et al.*, 2016; Sun *et al.*, 2017): extracting the appearance of features, deriving geometric features and utilising hybrid techniques. The geometric features of the face are obtained from elements of the face itself (the nose, the eyes, the eyebrows, the mouth, etc.) and face shapes. Meanwhile, appearance-based features are retrieved using the face's texture, wrinkles and any furrows that are caused by facial emotions. Deep learning using CNNs' architecture has become more popular throughout recent years because of its efficiency in extracting the features from image-based data, but it performs less well when the characteristics of the face are deformed. On the other hand, capsule networks can extract data from deformed images, and their intensive computation tasks can run on the graphics processing unit (GPU), which offers reliable results in a short amount of time.

3.2. Capsule Networks' Architecture:

This architecture created by Hinton *et al.* (2011) originated to replace CNNs' architecture. A capsule is a network of neutrons that accepts vectors as an input and an output, which is different from CNNs as they only accept scalar values. The capsule property of being equivariant gives it the ability to learn the deformations and the features of an image besides the viewing conditions. Then, each single capsule network contains a group of neurons in which their output represents a different feature for a similar characteristic. As a result, this advantage allows the system to recognise the entire face starting by recognising its elements. For example, when the CNN detects a face, it detects it even if it has an incorrect eye position. Therefore, equivariance makes sure that the features of the face are present and located in their natural position in the detected image. As a result, the efficiency of this property made it desirable for capsule networks.

There are three main methods for capsule implementations in the literature: capsules based on dynamic routing where each capsule can call active capsules from the levels below (Sabour *et al.*, 2017). Second, the capsule of transforming auto-encoders, which backpropagates the difference between target outputs and the actual ones, is used to learn the weights of the connections (Hinton *et al.*, 2011). Third, rather than utilising vector outputs, Hinton suggested that the input and output of a capsule should be represented as matrices. The dynamic routing was also replaced with an approach called expectation maximization (EM) to decrease the size of the transformation matrices among capsules (Hinton *et al.*, 2018). Capsules differentiate from classical CNNs that have been modified from scalar into vector features in capsules, and they use the dynamic routing method based on the same mechanism in place of the max-pooling layer. The use of max pooling was discontinued because it does not consider spatial relations and only retains prominent information, which makes the trained model incapable of recognising spatial positions between facial features. If the facial elements in an image are not organised in a natural order, CNNs still define them as faces; meanwhile, capsules work differently by using vectors to recognise faces out of order based on the spatial information that is already stored in the vector. This is the main disparity between the two methods.

Capsule networks and and a LightGBM have been applied in relation to facial emotion image classification during this specific project. Mapping a matrix of pixel values to an emotion classification involves levels of abstraction that make it applicable for studying deep architectures. More importantly, the learned intermediate representations of face types, given that the forehead, eyes and lips are part of the overall picture, are more qualitatively interpretable. The model efficiently masters this particular classification task once these features are learned in supervised training. Furthermore, this deep engineering outperforms results from more shallow networks.

3.3. Light Gradient Boosting Machine:

In 2017 Ke *et al.* (2017) introduced a new learning algorithm – the LightGBM. This approach is based on the platform of a gradient-boosting decision tree (Friedman, 2001). LightGBM is very accurate and displayed a fast level of training efficiency with regard to lots of applications as opposed to traditional gradient-boosting decision trees, which are considered time-consuming and have computational complexities. In addition, the LightGBM method has been used successfully for regression (Singh *et al.*, 2020) and classification (Yang and Shi, 2019). Its success gave the green light for new techniques to be used, including exclusive features bundling and one-sided gradient analysis.

A leaf-wise leaf development methodology with depth limitation has been used. Information about the positioning of the level may at the same moment divide the leaves of the same stratum, making multithreading optimisation simple and allowing complexity to be controlled. That will lessen the quantity of errors and will raise the degree of precision. Regarding the depth limitation of the leaf-wise, it can guarantee a high-productivity level as it is equipped to forestall the over-fitting simultaneously. The pace of the cache hit was streamlined, and the multithreading was upgraded. Meanwhile, LightGBM is a decision-making method that is built on a histogram-based decision tree and employs histogram subtraction. This method has added the standards of a decision to the features of the category to dodge extra computational and memory overheads. It is made through the transformation of features into a one-hot characteristic with multi-dimensions.

4. Experiments and Models

In this section, the outlines of the specific model, experiments and training steps are presented. The model of our approach to be learned was developed with the following function:

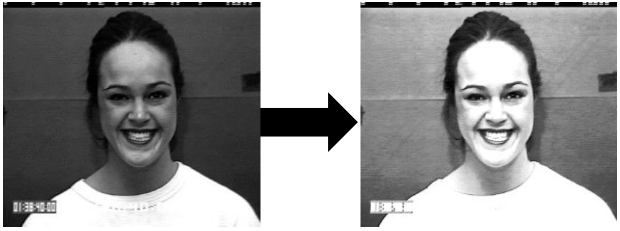
$$F\colon (X_j) \rightarrow \{P_i \mid 0 < i < 7\}$$

The input of the function is any given image (X_j), and it outputs the probability of belonging to each class (P_i). The main challenge faced here was using a small dataset. The equation is divided into two functions within the architecture because of the dual utilisations: a) identifying the features of the image and b) using those features for the classification. An important aspect of our approach is achieved through a deep neural network, and a probability prediction is realised using a LightGBM classifier, which is explained in subsections 4.2 and 4.3.

4.1. Pre-processing:

First, the presence of noise has an impact on the accuracy of FER. As a result, the pre-processing step within FER systems is extremely important and demands careful filter selection. In our work, the picture may have noise or blurring artifacts at first, which might degrade recognition accuracy. To increase the image's overall quality, the image-appearance filters were applied, specifically the median filter (MF) and the histogram-equalization (HE) technique on the original CK+ database to optimise the model's solidity in relation to noise. The median filter's function computes the median of the entirety of the pixels' values under the kernel window, and this result is used to replace the central pixel. This is highly effective for reducing random noise. The result is shown in Figure 1. Moreover, the HE technique was used to improve the picture's contrast and normalise its illumination effects to remove any abnormalities in the its lighting and background noise.

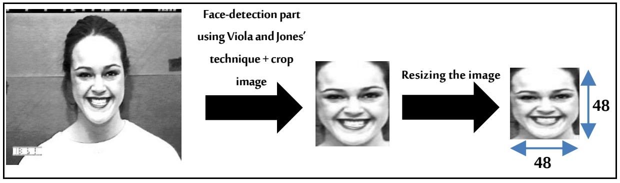
Figure 1: The Result of the First Step



Second, the employment of a face detector is a critical stage in the recognition process, since the face is the only important aspect of the picture, and all facial landmarks associated with emotions are only found on this part of the body. Because of the head's changing poses or movements, detection is often incorrect or too difficult, meaning the majority of face-detection techniques are based on the face alone. For Viola and Jones (2001), this technique was a well-known and extensively used classic face-detection method that was freely accessible in various modes of implementation; moreover, it is highly capable of detecting faces that are facing forward. Indeed, it is more reliable than other detection algorithms, such as the deep dense face detector (DDFD) method, and it also consumes fewer resources and takes less time. Raw pictures may be cropped to retrieve the facial region using the identified face-bounding box. This approach can minimise the amount of time spent on computational calculations, and it is able to highlight the actual facial area.

Finally, all cropped images were standardized to a uniform size of 48x48 pixels. This step is necessary to shorten the processing time, as shown in Figure 2.

Figure 2: The Result of Pre-Processing Step



4.2. Neural Network Architecture:

Neural networks are the de-facto standard for most image-processing tasks. Especially when it comes to feature extraction, recent advancements in CNNs and their spatial invariance have shown significant improvements in terms of accuracy and scalability. Nevertheless, the key drawback of such an approach is the need for a large dataset and long training times. To train a large neural network to an acceptable level of accuracy, thousands of images are needed, yet this was not the case during our research; a dataset of around 1,000 images was used. This makes learning which features to extract a significantly difficult problem. For this reason, in this research, the use of convolutional layers was limited. Instead, the capsule layers recently published by Sabour *et al.* (2017) were introduced.

Capsule networks try to deal with a theoretical problem introduced by convolutional networks. One main element that affects the success of convolutional networks is their max-pooling layers. This pooling operation is the reason for the loss of valuable spatial information between layers in convolutional networks. This drawback is addressed using a dynamic routing algorithm, which selects information and sends it to upper layers but only to capsules that really match those features. Capsule networks have displayed on-par performances with convolutional networks but with a smaller number of training parameters. This fact makes capsule networks a good option to train networks with small datasets; hence, in this study, the applicability of capsule networks combined with a classical approach was explored.

An overall summary of the layers used in the image-feature-identification network is shown in Table 1. Those layers can be categorised into three main components in the model, namely the convolutional module, the capsule module and the feature-explanation module.

Table 1: Composition of layers in neural network.

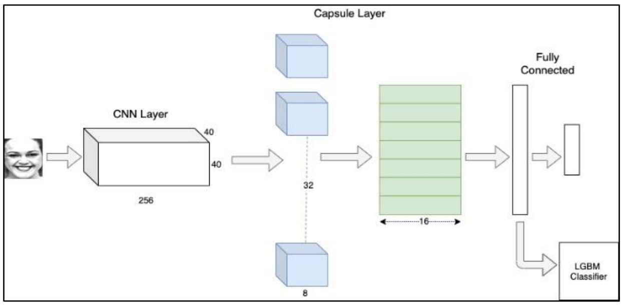
Layer	Units	Parameters
Convolutional	256 x (40x40) kernels	62464
Primary Capsule	32 8-dim capsules	5308672
Explain Capsule	7 16-dim capsules	7340032
Fully Connected 1	100 neurons	11300
Fully Connected 2	7 neurons	707

- **Convolutional module**: this module was used to expand the features of the image. In a deep convolutional network, convolutional layers initially learn very low-level image features like colour densities and edges. Here, the same was done using a convolutional layer.
- **Capsule module**: this is where the core functionality of the model lies. It has two layers; the layer of the primary capsule has 32 capsules, and each capsule has eight 9x9 kernels with a stride of two. There each capsule generates a 16x16x8 tensor, which is fed into the expansion layer that has seven capsules. Here, seven capsules were selected, which corresponds to the number of emotions in our classification.

> The output of the capsules was expected to be an expansive feature regarding how the attributes show up on the image effect for each class.

In order to optimise those predicted features, it was decided that a fully connected network would be utilised, which is known as a feature-explanation module. It has two layers with 100 and seven neurons, respectively. This whole deep neural network was trained as one separate network to perform the emotion classification task. Our deep capsule network architecture (Figure 3) has three main components: the convolutional layer, the capsule module, and the fully connected feature-expansion component. The representation of extracted features from the first fully connected layer was fed into the LightGBM classifier for further classification.

Figure 3: Overall Architecture of our Model



4.3. Classical Classifier:

The accuracy achieved when using the deep neural networks in isolation could be improved by employing an ensemble model; an ensemble would be created using a classical classifier. The utilisation of classical approaches combined with a methodology like SVM is standard. Nevertheless, a classical classifier that was combined with the features extracted from a deep neural network was used here. During this research, several classifiers were tested but we finally settled on an LGBM classifier with 70 estimators. LGBM classifiers use Light GBM algorithm, which in turn utilise algorithms relating to tree-based learning. The novelty of the LightGBM is that it grows trees vertically, while other algorithms do it horizontally. Since, it grows tree leaf-wise, it will then grow the leaf with the greatest delta loss, which reduces the occurrence of even greater losses compared to level-wise algorithms.

4.4. Optimisation and Learning:

The learning within this model is a two-part process. First, it needs to optimise and train the deep neural networks and then train the classical classifier with selected features that have selected dimensions. When training the neural network, it was done as an end-to-end process. The final target of the network is not feature engineering but doing a final prediction for detected emotion. It must optimize a cost function established with the intention in order to learn optimal settings. The categorical cross-entropy function is used to measure prediction accuracy:

$$L(X_i, \theta) = -\sum_{i}^{C} t_i \log F(X_i)$$

Where

X_i: input image

θ : parameters of the model

C: number of classes

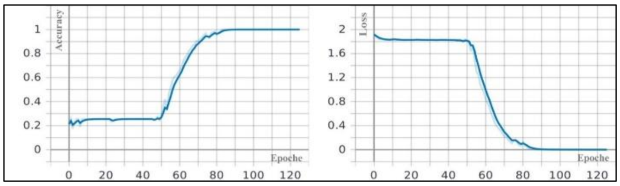
t_i: the expected prediction

The output is one-hot encoded; hence, t_i is non-zero for only one class. The above equation tries to maximise the probability of that expected class. Moreover, since the probability of other classes is ignored, it has to use softmax activation in the last layer of the model. This makes sure the sum of all the probabilities equals one. After the deep learning model is trained to a sufficient level, an accuracy rate of about 91% using this model alone is achieved. Nevertheless, with an ensemble model, the improvement in accuracy could be enhanced. Before the final classification layer is completed, the feature representation for the image from the dense layer as an input to the LightGBM classifier is used. LightGBM is a classifier that is trained with an input vector of 100 dimensions that are compared with the class labels.

4.5. Training and Dataset:

Our models are trained on a machine with Windows 10 Professional, an Intel Xeon processor and a NVidia K80 GPU with a 24GB memory. As shown in Figure 4, the model was trained for 100 epochs; during this time, training accuracy grew from 20% to 97%. Also, as observed after 80 epochs, the model started to overfit, meaning it was suitable to use for the evaluations.

Figure 4: Loss and Training Accuracy of our Model



At the same time, the CK+ (Lucey *et al.*, 2010) dataset was also used to train our dataset, which contains images belonging to the seven categories of emotions.[1] Altogether, there 984 images. The dataset with a 0.7 ratio to the validation and training set was split. Owing to the database's dimensions, the validation set for testing was used as well. Different classes of dataset did not have the same number of images. Specifically, the image distribution was 135 image showing anger, 207 images of happiness, 249 images denoting surprise, 75 images showcasing fear, 84 images displaying sadness, 177 images highlighting disgust and 57 images showing contempt.

5. Results and Discussion

In this research paper, our main model consists of a capsule-based feature extractor and a LightGBM classifier. The primary model obtained a 91% accuracy rate during the test with the CK+ dataset, as shown in Table 2.

Table 2: Accuracies Achieved for Test Set with Different Models

Model	Accuracy
CN- based model	0.814
SVM classifier	0.746
Capsule-based model	0.854
Capsule feature extractor + LightGBM classifier	0.910

Apart from this main model, a few other experiments were conducted in order to compare our approach with others, including a classical CNN-based image classifier and an SVM classifier. In addition, before integrating the LGBM classifier, only the capsule-based model was used as a benchmark for the expected results. Table 3 depicts the results achieved through these different approaches.

Table 3: Matrix Confusion of our Approach

Emotions	Angry	Contempt	Disgust	Happy	Fear	Surprise	Sad
Angry	0.9062	0.0000	0.0000	0.0000	0.0000	0.0000	0.0938
Contempt	0.0000	0.7391	0.0000	0.0000	0.1304	0.0000	0.1304
Disgust	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Happy	0.0000	0.0182	0.0000	0.9455	0.0364	0.0000	0.0000
Fear	0.1200	0.0400	0.0000	0.0000	0.8400	0.0000	0.0000
Surprise	0.0000	0.0137	0.0000	0.0000	0.0548	0.8904	0.0411
Sad	0.0870	0.0000	0.0000	0.0000	0.0000	0.0000	0.9130
Accuracy	0.9061 = 90.61 %						

[1] https://sites.pitt.edu/~emotion/ck-spread.htm (with permission from copy right holder (©Jeffrey Cohn))

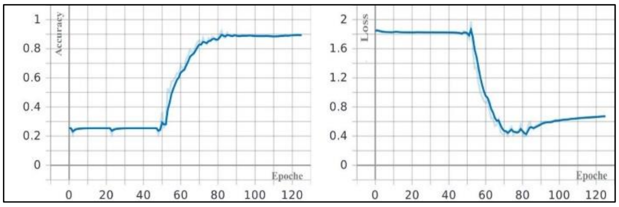
When the results of Table 3 are examined, a few important facts about the model can be noticed. Compared to other models, our proposed architecture has achieved very high levels of accuracy. This high accuracy rate can be attributed to two separate modules: the capsule feature extractor and the LightGBM classifier. Other baseline methods were conducted in order to get an understanding of this contribution.

Usually, deep learning methods are very good at image classification and feature extraction, especially CNNs. Nevertheless, in our case, the convolutional layer-based method was only able to achieve 81.4% accuracy. However, the model capsule layer-based was able to achieve 85.4% accuracy. This observation can be attributed to the size of the dataset that was used; our training dataset only had around 600 images as capsule networks have a superior ability to converge using only a very small number of samples. That is because there is no data loss as in convolutional neural layers. Since CNN uses max-pooling, which literally selects and keeps the most significant number in a matrix, which contrasts with the dynamic routing mechanism used in capsule layers, there could be high chance of data loss in convolutional layer-based models.

In our research, a capsule network was only used for the purpose of feature extraction rather than classification. In order to make sure that this is the optimal method for extracting features, a SVM classifier was used and its results were compared its with those from the capsule-based model. The SVM-based model was only able to achieve 74.6% accuracy. This is because it does not have the advanced capability of capsule networks or convolutional networks to extract important features. Instead, SVM tries to use all of the features that are fed into the classifier model.

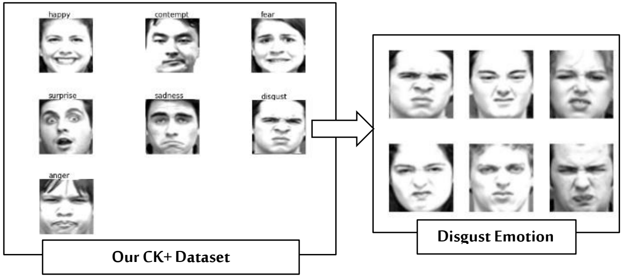
As a final step in our experiment, our capsule model was integrated with an LightGBM classifier and was able to achieve around a 6% improvement in accuracy. This means it can be hypothesised that capsule networks are better feature extractors and the LGBM classifier can conduct better classifications when the dataset is small. The accuracies achieved with the final capsule model are not similar throughout all the six classes in the database. Below, Figure 5 represents our obtained results of the proposed approach.

Figure 5: Loss and Accuracy of our Model Validation



Only the emotion class of disgust achieved 100% accuracy. This could be because significant visual cues are available in this category of pictures. Figure.6 highlights some samples from the disgust emotion category.

Figure 6: Our CK+ Dataset with Samples from Disgust Category.



In all the samples of the disgust category in Figure 6, there is a main common visual cue around the eyes. Among these faces, the eyes are shown as a continuous black shape; this feature might be the cause of there being such a high level of precision. In any of the other classes, such noteworthy features were not observed.

6. Conclusions

A novel technique for face emotion identification is provided in this research. The suggested system can extract feature points and recognise facial displays of emotion from pictures. However, the extraction of exact facial features may be a difficult process at times, and it generally necessitates a large number of calculations. Using capsule nets for the extraction of characterisations was proposed in our study, followed by the employment of LightGBM for the classification process. The average identification performance for facial expressions might be as high as 91% accurate. In comparison with some current methods, this outcome is highly promising.

There are some ideas and many different experiments, tests adaptations that have been left open for future research. Upcoming work will concern itself with applying this model to other datasets (i.e. big datasets, multi-operation datasets and sequence videos). In addition, different methods using three-dimensional models will be tested and will integrate a new category of emotions.

Biographies

Tegani Salem

Department of Electrical Engineering, Faculty of Science and Technology, Biskra University, Biskra, Algeria, salem.tegani@univ-biskra.dz , 00213663544420

Mr. Tegani is an Algerian PhD student in electrical engineering. He obtained a master's degree in computer science (Decision-making and Multimedia) from Biskra University in Algeria in 2016. He is also a member of the Expert Systems, Imaging and their Applications in Engineering (LESIA) Research Laboratory at Biskra University and the PRFU project (Big Data Classification Using Machine Learning Approaches. Code: C00L07UN070120220004). He is currently working on a doctoral LMD thesis – *Recognizing Facial Emotions Using Deep Learning and Multiple-observations Datasets* – and specialises in remote monitoring.

Telli Abdelmoutia

Department of Computer Science, Faculty of Exact Sciences and Natural and Life Sciences, Biskra University, Biskra, Algeria, a.telli@univ-biskra.dz, 00213541340479

Dr. Telli is an Algerian Lecturer. He received his PhD from Biskra University in Algeria in 2018 with a cotutelle from Artois University in Lille (France). Additionally, he received his habilitation of science from Biskra University in 2021. His current research interests include the representation of knowledge, big data, machine learning, the management of uncertainties and conflicts in priority knowledge bases. Furthermore, he is the Head of the PRFU project. ORCID: 0000-0002-2907-9782.

References

Black, M.J. and Yacoob, Y. (1995). Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In: *Proceedings of IEEE International Conference on Computer Vision*, Cambridge, MA, USA, 20-23/06/1995. DOI: 10.1109/ICCV.1995.466915.

Corneanu, C.A., Simón, M.O., Cohn, J.F. and Guerrero, S.E. (2016). Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**(8), 1548–68. DOI: 10.1109/TPAMI.2016.2515606.

Darwin, C. and Prodger, P. (1998). *The Expression of the Emotions in Man* *and Animals*. Oxford: Oxford University Press.

Ekman, P. (1993). Facial expression and emotion. *American psychologist,* **48**(4), 384–92. DOI : 10.1037/0003-066X.48.4.384.

Essa, I.A. and Pentland, A.P. (1997). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **19**(7), 757–63. DOI: 10.1109/34.598232.

Friedman, J.H. (2001). 1999 Reitz Lecture. *The Annals of Statistics,* **29**(5), 1189–232.

Friesen, W.V. and Ekman, P. (1976). *Pictures of Facial Affect*. Palo Alto, CA, USA: Consulting Psychologists Press.

Hinton, G.E., Krizhevsky, A. and Wang, S.D. (2011). Transforming auto-encoders. In: *International Conference on Artificial Neural Networks*, Espoo, Finland, 14–17/06/2011.

Hinton, G.E., Sabour, S. and Frosst, N. (2018). Matrix capsules with EM routing. In: *6th International Conference on Learning Representations*, Vancouver Convention Center, Vancouver, BC, Canada, 30–03/04–05/2018.

Hong, C., Chen, L., Liang, Y. and Zeng, Z. (2021). Stacked capsule graph autoencoders for geometry-aware 3D head pose estimation. *Computer Vision and Image Understanding*, **208**(n/a), 103224.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W. and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, **30**(n/a), 3146–54.

Kim, B.-K., Lee, H., Roh, J. and Lee, S.-Y. (2015). Hierarchical committee of deep cnns with exponentially-weighted decision fusion for static facial expression recognition. In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Seattle, Washington, USA, 09–13/11/2015.

Kim, S., Kavuri, S. and Lee, M. (2013). Deep network with support vector machines. In: *International Conference on Neural Information Processing*, (pp. 458–65), Springer, Berlin, Heidelberg, 3–7/11/2013. DOI: 10.1007/978-3-642-42054-2_57.

Ko, B.C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors,* **18**(2), 401. DOI : 10.3390/s18020401.

Liu, M., Shan, S., Wang, R. and Chen, X. (2014). Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28/06/2014.

Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, San Francisco, CA, USA, 13–18/06/2010.

Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion,* **16**(4), 363–8.

Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences,* **2**(3), 1–8.

Mellouk, W. and Handouzi, W. (2020). Facial emotion recognition using deep learning: Review and insights. *Procedia Computer Science,* **175**(n/a), 689–94.

Minaee, S., Minaei, M. and Abdolrashidi, A. (2021). Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors,* **21**(9), 3046. DOI: 10.3390/s21093046

Patrick, M.K., Adekoya, A.F., Mighty, A.A. and Edward, B.Y. (2019). Capsule networks – a survey. *Journal of King Saud University, Computer and Information Sciences,* **x**(x), x-x. DOI: 10.1016/j.jksuci.2019.09.014

Ranzato, M.A., Susskind, J., Mnih, V. and Hinton, G. (2011). On deep generative models with applications to recognition. In *CVPR* 2011, **n/a**(n/a), 2857–64.

Sabour, S., Frosst, N. and Hinton, G.E. (2017). Dynamic routing between capsules. In: *Proceedings of the 31st International Conference on Advances in Neural Information Processing Systems*, Long Beach, California, USA, 04/12/2017.

Singh, S.P., Singh, P. and Mishra, A. (2020). Predicting potential applicants for any private college using lightGBM. In: *2020 International Conference on Innovative Trends in Information Technology (ICITIIT),* Kottayam, India, 13–14/02/2020.

Sun, W., Zhao, H. and Jin, Z. (2017). An efficient unconstrained facial expression recognition algorithm based on stack binarized auto-encoders and binarized neural networks. *Neurocomputing,* **267**(n/a), 385–95.

Tereikovska, L., Tereikovskyi, I., Mussiraliyeva, S., Akhmed, G., Beketova, A. and Sambetbayeva, A. (2019). Recognition of emotions by facial geometry using a capsule neural network. *International Journal of Civil Engineering and Technology,* **10**(3), 1424–34.

Terzopoulos, D. and Waters, K. (1990). Analysis of facial images using physical and anatomical models. In: *Proceedings Third International Conference on Computer Vision*, Osaka, Japan, 4–7/12/1990.

Tiwari, S. and Jain, A. (2021). Convolutional capsule network for COVID-19 detection using radiography images. *International Journal of Imaging Systems and Technology,* **31**(2), 525–39.

Valdenegro-Toro, M., Arriaga, O. and Plöger, P. (2019). Real-time convolutional neural networks for emotion and gender classification. In: *Proceedings of ESANN 2019 Conference, European Symposium on Artificial Neural Networks*, Bruges, Belgium, 24–26/04/2019.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001, Kauai, HI, USA, 8–14/12/2001.

Yacoob, Y. and Davis, L.S. (1996). Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **18**(6), 636–42.

Yang, C. and Shi, Z. (2019). Research in breast cancer imaging diagnosis based on regularized lightGBM. In: *International Conference on Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health*, Beijing, China, 16–18/12/2019

Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y. and Dobaie, A.M. (2018). Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing,* **273**(c), 643–49.

Zhang, J. and Xiao, N. (2020). Capsule network-based facial expression recognition method for a humanoid robot. *Recent Trends in Intelligent Computing, Communication and Devices*, **1006**(n/a), 113–21.