

# Computer Vision: The Effectiveness of Deep Learning for Emotion Detection in Marketing Campaigns

Shaldon Wade Naidoo, Nalindren Naicker\*, Sulaiman Saleem Patel and Prinavin Govender

Department of Information Systems,  
Durban University of Technology,  
Durban, South Africa

**Abstract**—As businesses move towards more customer-centric business models, marketing functions are becoming increasingly interested in gathering natural, unbiased feedback from customers. This has led to increased interest in computer vision studies into emotion recognition from facial features, for application in marketing contexts. This research study was conducted using the publicly-available Facial Emotion Recognition 2013 data-set, published on Kaggle. This article provides a comparative study of four deep learning algorithms for computer vision application in emotion recognition, namely, Convolution Neural Network (CNN), Multilayer Perceptron (MLP), Recurring Neural Network (RNN), Generative Adversarial Networks (GAN) and Long Short-Term Memory (LSTM) models. Comparisons between these models were done quantitatively using the metrics of accuracy, precision, recall and f1-score; as well and qualitatively by determining goodness-of-fit and learning rate from accuracy and loss curves. The results of the study show that the CNN, GAN and MLP models surpassed the data, and the LSTM model failed to learn at all. Only the RNN adequately learnt from the data. The RNN was found to exhibit a low learning rate, and the computational intensiveness of training the model resulted in a premature termination of the training process. However, the model still achieved a test accuracy of up to 72%, the highest of all models studied, and it is possible that this could be increased through further training. The RNN also had the best F1-score (0.70), precision (0.73) and recall (0.73) of all models studied.

**Keywords**—computer vision; deep learning; emotion detection; generative adversarial networks, marketing campaigns component

## I. INTRODUCTION

The evolution of technology has enabled businesses to gain a better understanding of their customers and develop products and services to better accommodate their markets [1]. The growth of customer centric business models has led to a variety of research studies in business and academia. Notably, Zaim et al in [1] have studied different analytical models for measuring customer experience (CX). A common requirement of these analytical models is that subjective opinions on CX are needed. These opinions are typically captured using surveys and similar methods in [2] and [3]. However, these methods are subject to bias due to human elements such as dishonesty and pressure. This has motivated for more technologically driven, automated approaches to soliciting customer feedback which mitigate some of these biases. One such approach is the use of computer vision to recognise human emotion from facial features. In these studies, cameras were used to record the facial expressions of individuals in real-time. The video

footage was then processed to determine emotions based on the facial expressions that were recorded, and hence obtained a more authentic source of CX feedback.

In this work, we evaluate and compare a variety of machine learning algorithms that perform human emotion recognition in the context of marketing campaigns. Our study focuses on the following three: happiness, anger and surprise. In the context of marketing studies, happiness is an emotion that indicates that adverts are received favourably, which strengthens brand reputation [2]. Anger is a powerful emotion that can evoke responses from communities, and can be leveraged in social marketing campaigns around sensitive or political issues [2]. Finally, the emotion of surprise creates a sense of urgency in a marketing campaign, and may result in consumers taking more immediate action (e.g. online popup advertisements that are valid for a limited time) [2]. Emotion recognition is a classification problem, which literature indicates can be solved using the “deep learning” class of machine learning algorithms [4], [5]. Deep learning is a class of machine learning algorithms that mimics the way the human brain functions, and are typically used in solving prediction or classification problems [6]. The review study by Liang et al. in [7] showed that the convolutional neural network (CNN) deep learning algorithm is popular when performing facial expression recognition (FER). In this paper, other deep learning algorithms were explored and compared to the CNN model to perform FER; viz. the Multilayer Perceptron (MLP), Generative Adversarial Networks (GAN), Long-Short term memory (LSTM) and Recurring Neural Network (RNN) models were studied. The performance of these models was evaluated using accuracy, f1-score, precision, and recall as metrics [8].

May 24, 2022

## II. LITERATURE REVIEW

In their 2017 study, Siau and Yang in [6] stated that marketing is a field that has been impacted by advanced technology. They elaborated that the marketing field is the most vulnerable to be radicalized by the Fourth Industrial Revolution. In the years following the article, there were significant findings that were made in which the above predictions were fulfilled. Articles that prove this are discussed in the literature review that follows.

Ładyżyński, Źbikowski in [9] experimented in 2019 on a system that directs marketing campaigns to targeted customers,

utilising machine learning techniques. This paper proposed the use of several machine learning algorithms such as Random Forest classifiers, Deep Belief Networks and Classification and regression (CART) classifiers. The team made use of a time series approach. In addition, in order to get results, a marketing campaign simulator was used. This simulator mimicked the call centre scenario described in the paper. Metrics such as precision and recall were applied to evaluate the models.

In the year 2020, Koehn, Lessmann in [10] published findings in which predictions were made about the online shopping behaviour from click-stream data using deep learning. In the paper, the data of user sessions were utilised and mined within a time frame of three months. The data was submitted into various classification models such as Multilayer Perceptron, Long Short-Term Memory (LSTM), Gradient Boosting and Gated Recurrent Networks. The results from this model were based on revenue, the amount of revenue each correct prediction brought in and model accuracy.

Cheng and Tsai in [11] published results in 2019, in which three deep learning models were applied for automated sentiment analysis using social media data. The use of WOM (word of mouth data) was supplied into a LSTM, Bidirectional LSTM (BiLSTM) and Gated Recurrent (GRU) model. The methodology consisted of employing a web crawler in order to gain the data. The data underwent pre-processing due to anomalies such as colloquialism and emotions. The data was labelled using tools such as NTLK and MS text block API, embedded through GLoVe and eventually distributed into the three models. The models were equipped to predict sentiment, and were evaluated using metrics such as precision, recall, F-measure and accuracy [11].

Reviewing these four articles and the respective methodologies proved that Siau and Yang's findings in [11] that was published in their article, which suggests technology is taking over the marketing field at an alarming rate. A common topic among these four articles was the use of Deep Learning. Deep learning is the ability to mimic behaviours of the human brain to solve problems using technology. Emotion depicts a person's feelings about a certain situation and is often portrayed by a facial expression. Therefore, if we were to detect a persons' facial expression, one can classify how a person ought to feel about certain scenarios. In the next section of our literature review, the relevant findings on the use of deep learning in emotion detection are going to be explored further.

Ko in [12] published a brief review of facial emotion recognition methods. This paper stated that there are two types of ways in which facial recognition can be detected using static and dynamic images. Further on into the paper, the process of recognizing these emotions were described. The process flows from the input images, followed by facial detection and landmark detection, feature extraction and lastly emotion classification were discussed. This process is a standard for all facial emotion recognition experiments; however, it may be flawed. Due to the process the raw image itself is not used and only features are extracted. There are many different models in which can be used to detect facial emotion recognition.

The use of a CNN model for Facial emotion recognition (FER) is common. Tümen, Söylemez in [13] used a CNN model in their paper in which had a 57% success rate. The

methodology included using images from the FER 2013 data-set. Subsequently, the images were supplied into a 3-layer CNN model after feature extraction. The draw-back from this experiment was the data split. Tümen, Söylemez in [13] split the FER data-set into an 80% for training 10% for testing and 10% for validation.

An MLP model is a multi-layer perceptron which consists of multiple neural layers. This model, as used by Tarnowski, Kołodziej in [4] in their findings, is the second most common FER model. Tarnowski, Kołodziej in [4] published findings in which real time 3D data was used to classify amongst the MLP model. The input consisted of a data size of 12 men, in which is relatively small and biased. However, the team was able to obtain a model accuracy of 73%. The methodology used included using a Microsoft Kinect camera in which 12 men sat two metres away from each other. Each participant was tasked with mimicking certain facial expressions. The data collected was deposited into the model, which consisted of seven neural layers, and finally a classification was done. The use of live data left room for many flaws. The response time of the participant to mimic expressions, the lighting conditions, The limitations of the Kinect camera and the size and biases of the data-set were a part of several flaws in this title.

GAN models are an interesting topic. Yi, Sun in [5] experimented on image augmentation using a GAN model. Images were used from the FER 2013 data-set in which a GAN model was able to augment some images and used the images normally. In the paper, it was found that a higher accuracy was generated by using the GAN for image augmentation. However, in this paper, a CNN was done to generate these results as opposed to a GAN model. These models do have potential in facial recognition as stated by Luo, Zhu in [14].

Another deep learning model which is overlooked due to CNNs are LSTM-RNN models. Sepas-Moghaddam, Etemad in [15] stated in their paper that the tested method of an LSTM-RNN model proved to be superior to the normal CNN way of image classification. In the publication, the use of a CNN VGG16 model extracted special features and also added an attention mechanism, which enables effective learning, before running the image through a Bi-directional LSTM-RNN model. The experiment used 800 images from the LFFD data-set. In the absence of the attention layer, the normal LSTM model produced an accuracy of 80%. There is room for improvement with this model. A bigger data-set with more expressions is recommended to solidify the results.

The manner in which technology has impacted the marketing field and how deep learning has been used to identify facial expressions and classify it as an emotion has been expressed in aforementioned statements. Reyes in [16] stated in an article that expression can be derived from emotion, and that emotion is a type of universal language, which stands in the gap for verbal communication. This leads us onto the last section of this literature review. The following section will discuss emotion detection in Marketing.

Yolcu et al in [17] published findings about a deep learning face analysis system to monitor customer interest. In their findings they monitored customers head position and facial expression to determine how interested customers are. In the methodology of the article Yolcu et al, used the Viola and

Jones algorithm for feature extraction, in which the image was cropped and only the head pose, and facial expression was extracted. A 3-layer CNN model was used to classify 8040 images for head position and 1206 for facial expression, from the Radbound Face Database. To verify this system the KDEF database was used with 4900 images and seven different emotions. A 90:10 split was done for model training and testing. A 94% accuracy was achieved for emotion accuracy in this article. Yolcu et al, used a mask generation flow in order to achieve such high results. The images were also pre-processed twice, using the Viola and Jones algorithm and also another CNN model. To improve this research, a multimodal approach in which multiple models can be used are suggested.

Ceccacci, Generosi in [3] experimented on a deep learning system which tracked and monitored customer behaviour in store. The system conducted testing in real life, using a Logitech Quickcam 4K camera. This experiment input the images of customers into these cameras which was linked to a CNN model. A total of 30 customers, split equally of gender, were tested. Images of them were distributed into a CNN model that was trained and tested using FER+ and EmotioNet data-sets in which consists of millions of images between the two. Model accuracy was not stated in the article. However, the model was able to predict 66% of emotions in the live experiment. These are two of the articles in which face emotion recognition was processed and completed using deep learning.

In our review of similar work, we have discussed deep learning in marketing, the use of deep learning in emotion detection and emotion detection in marketing. It can be assumed that there are gaps in literature that can add to how one can utilise deep learning in emotion recognition in order to aid marketing strategies.

### III. METHODS AND MATERIALS

#### A. Data Acquisition

This study uses the Facial Emotion Recognition 2013 (FER 2013) data-set, available publicly on Kaggle [16]. The FER 2013 data-set consists of 35 685 greyscale images of size 48x48 pixels, and the provides recommended splits for testing and training image sets.



Fig. 1. Sample Images: Angry

In our study, the original FER 2013 data-set was first divided into two smaller data-sets. This was a result of limited processing power available by the HP i5 4GB RAM that was

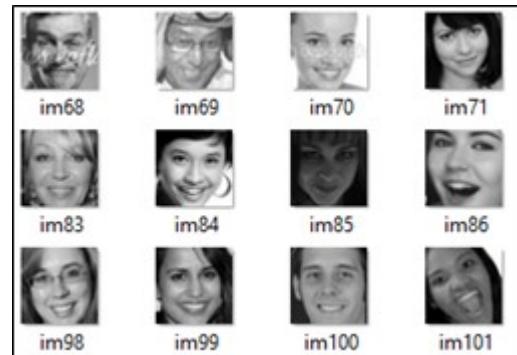


Fig. 2. Sample Images: Happy



Fig. 3. Sample Images: Surprised

used in performing these experiments. Both data-sets were then subdivided into training and testing sets, using an 80:20 ratio. In this research experiment, three emotions from the original FER 2013 data-set were considered (i.e. happy, angry, and surprised). Sample images from the FER 2013 data-set are provided in Fig. 1, 2 and 3, which show sample images of angry, happy and surprised facial expressions, respectively. The motivation for selecting these three emotions for the scope of this study was given in Section I.

The first data-set contained 12 000 images from these three images classes, consisting of 3 000 training images and 1 000 test images per emotion class. A similar structure was used for the second data-set, in order to avoid any type of bias. This data-set consisted of 21 000 images (5 600 training images and 1 400 testing images per emotion class).

#### B. Data Pre-Processing

There are generally two approaches to performing facial expression recognition (FER) using computer vision. The first is to pre-process images and extract features from them, which form the input to the machine learning algorithm. This has been shown to provide robust machine learning models in some studies, for example the study by Ceccacci et al [3]. The second approach is to use raw pixel data directly, which allow deep learning models to identify features for themselves during the model training process [18]. Exploratory checks performed on the data indicated that quality of images was high, and hence the experiment was designed using raw pixel data with no feature extraction or cleaning. However, model-

TABLE I. EXPERIMENT PROCESS

Step No.	Step Name	Details
1	Imports	Relevant import classes were imported into the experiment. Consistent imports namely would be Tensorflow, Keras, Matplotlib.pyplot, OpenCV and OS.
2	Pre-processing	A method is created in which sorts data according to its label and expression and populates them into an array. Images are also resized in this method
3	Train and Test variable population	The above method creates populates the relevant data into their train and test variables.
4	Array Reshape	The array size of each image is reshaped according to the model requirements.
5	Hyper-parameters set	A model is called in which hyper parameters are set. All models consist of the same hyper-parameters, except for layers and epochs. This is due to model requirements.
6	Model compiled	The model is compiled and set to produce model accuracy. Categorical cross-entropy to help us measure the performance of the classification output and tell us the difference between the classifications.
7	Epochs set, model fitted.	Epoch numbers are set and the model is fitted with training and testing data. The output of this step would be the accuracy of the model.
8	Graphs and Table of results.	Graphs are created to measure validation loss and accuracy. Table of results calculated

specific data transformations were necessary (discussed further in subsection 4).

According to [19], data segmentation and data transformation are critical factors in data pre-processing. Consequently, in this experiment, data segmentation was performed first by splitting the training and testing data to ensure we kept the integrity of the 80:20 split. Furthermore, each emotion had been allocated its own files within the testing and training folders. Each image was resized using OpenCV, to ensure the data would meet model requirements. In order to ensure the deep learning model was able to learn properly, each image was converted into an array. In which, the label and feature of each image was stored respectively.

Singh and Singh in [20] stated in their article that normalization of data in deep learning is imperative in order to make a good contribution to each feature. The writers also stated that normalization is a critical success factor in the learning of each algorithm. Therefore, in our experiment, we normalize the data according to the requirement of the algorithm being used.

Concluding our data pre-processing, the use of methods such as data segmentation, data transformation and data normalization ensured that each model that was used has an equal chance at performing its best.

### C. Design of Study

The experimental design method was used in the study of the effectiveness of deep learning for emotion detection in marketing campaigns. This type of study proved effective in a number of deep learning articles [4], [21], [22]. The independent variable being the different types of deep learning models such as CNN, GAN, MLP, RNN and LSM. The dependent variable was our metrics such as model accuracy, f1-score, precision and recall. The model accuracies depend on the type of model in order to prove the title of our paper. Experiments carried out on both data-sets followed the same steps and procedures, to prevent biasness in any way. Our experiments were conducted on Jupyter Notebook, using Python 3.0. It consisted of eight steps, as indicated in Table I above.

The steps shown Table I were a general guideline of how we carried out the experiment on how effective deep learning

was used for emotion recognition. The process outlined is consistent with the extant literature, such as in [23].

### D. Algorithms

In the study that was conducted, we focused on five deep learning models. Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Multilayer Perceptron (MLP), Generative Adversarial Network (GAN) and Recurrent Neural Networks (RNN). In this subset of our paper, we will be discussing in depth, these models and their unique capabilities they bring to the deep learning field. We will also discuss how they have been implemented and define functions used to generate appropriate outputs.

*1) Convolutional Neural Network (CNN):* CNN classifiers have been getting high amount of recognition in the deep learning world [24]. Most image classification problems use CNN models, due to the high levels of model accuracy, CNN has also proved to be the better and most perred model amongst most deep learning models for image classification [25]. Convolutional neural networks generally consist of two layers. The first being a convolution layer, also known as a C layer. The second layer being the subsampling layer, known as the S layer. Each S layer follows a C layer as depicted in graph below. Convolutional Neural Networks have an advantage over normal deep learning models, having the ability to accept 2D images without major changes to the array.

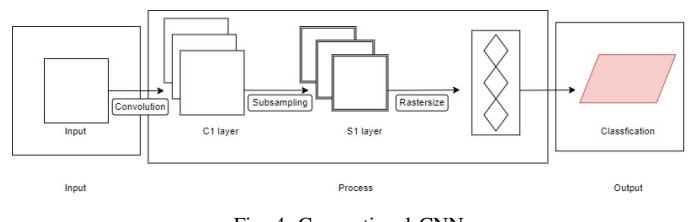


Fig. 4. Conventional CNN

The illustration depicted in Fig. 4 is a basic CNN model. Normally CNN models consist of two C1 layers and two S1 layers. However, to explain its structure we will use a single C1 and S1 layer. The input image, usually in the form of an array, would be fed into the first C1 layer. In this layer, feature maps would be formed through the convolutions. These feature maps would be inputted next into the S1 layer. During sampling the

feature map size would be reduced by a pooling method which normally consists of  $2 \times 2$ . Usually this process is repeated, depending on the number of layers. After the S1 layer the data is rasterized and a classification is formed.

The CNN model that we experimented on consisted of five convolution layers, in which had a kernel size of 3. Five Subsampling layers of pooling size  $2 \times 2$ . We used the Rectified Linear Unit (ReLU) activation function to output the input only if the output is a positive value. We used batch normalization to improve the speed of our model and finally a dropout of 0.20 is used in order to prevent overfitting. The equation for the ReLU activation function is described as:

$$f(x) = \max(0, x) \quad (1)$$

**2) Multi-Layer Perceptron (MLP):** MLP models are binary classifiers in the field of deep learning. General MLP models consist of multilayers however, single layers are also used. MLP models are famously used to state whether an input is something or not. Each Perceptron model consists of three layers. An input layer, a hidden and an output layer. Perceptron has a general rule, it states that the model will learn the best weight coefficients.

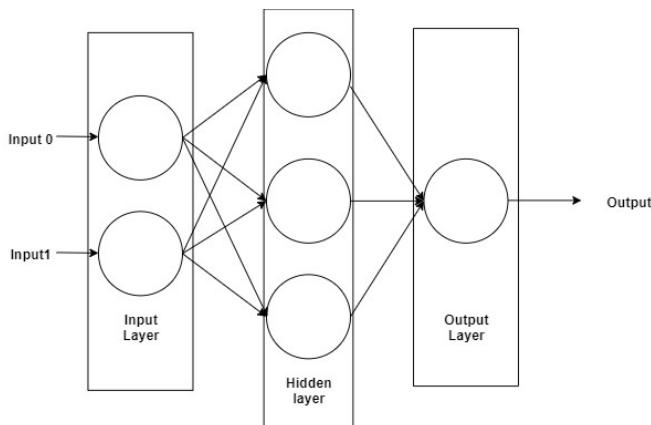


Fig. 5. MLP Model

Following from the beginning of the diagram in Fig. 5 the inputs were read in the input layer. In the hidden layer the weights were then calculated and their net input function was determined. After this, their activation function transformed the net input into an output, within the output layer. If any errors occurred, they were caught and sent back to the input layer. The perception activation function is described as:

$$f(x) = \begin{cases} 1 & \text{if } \mathbf{w}\mathbf{X} + \mathbf{b} > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where  $\mathbf{w}$  is the real weight value matrix,  $\mathbf{b}$  is the bias vector and  $\mathbf{X} = [x_1 \dots x_k]^T$  is the  $k$ -dimensional vector of input data. Note that the boldface in (2) represents matrix variables, rather than scalar variables and that  $(\cdot)^T$  represents the matrix transpose.

As stated above, each MLP model consists of three layers. In the research that we have carried out, the MLP model that we used consists of two dense layers. The first dense layer

having 128 neurons and using the ReLU activation function. The second layer consisting of 5 neurons and the softmax activation function

#### E. Generative Adversarial Network (GAN)

The popularity around GAN models stems around their ability to augment data. GAN models are used famously throughout the deep learning fields to enhance other deep learning algorithms [26]. GANs are generally made up of two smaller deep neural networks. The first being a generator, in which is responsible for generating data. The second being a discriminator, which takes the real and fake images to classify which one is real or fake.

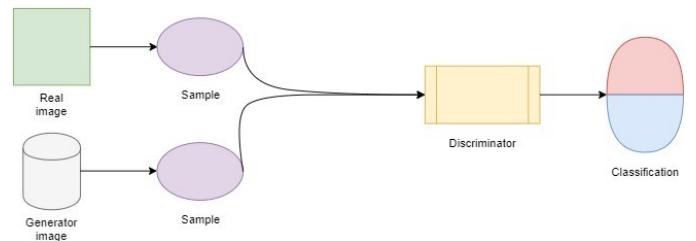


Fig. 6. GAN Model

Fig. 6 depicts the general structure of a GAN model. In the experiment that we have conducted, the GAN model does not consist of a generator, due to the input images already being available. However, the discriminator that we have used consists of a method created. The discriminator consists of two down sample layers consisting of three kernels each and LeakyReLU as an activation function, which was suggested from reviewing the works of Krestinskaya, Choubey in [27]. The activation function for LeakyReLU is described below.

$$f(x) = \max(0.01x, x), \quad (3)$$

where  $x$  is the input to the activation function.

Our discriminator also has a flatten feature, followed by a 0.4 dropout layer. The output layer of our GAN model has a softmax activation function. It is understood that GAN models classify whether an image is fake or not. However, in our research we were able to recognize whether an emotion was present or not by using three GAN models developed independently. In this study we utilized the GAN models to independently test for the emotions “happy”, “angry” and “surprised”. The structure of the GAN was based on the work in [28].

#### F. Recurring Neural Network (RNN)

The operations and structures of RNNs are the same as Fig. 7. RNNs feed the information through their vectors back into their input gates. In the RNN model that we used, the hidden layer consisted of 64 neurons, a dropout layer and a recurrent dropout layer. In the second hidden layer we used the ReLU activation function. For the output layer we use a dropout of 0.5 and the softmax activation function.

#### G. Long Short-Term Memory (LSTM)

LSTM models are a type of RNN deep learning model. Conventional neural networks have a feed forward tendency. However, the LSTM models have a feedback connection. LSTM neural networks are advancements of the general Recurrent neural network model and are known for modelling chronological sequences [29]. However, they can be used to classify images [30].

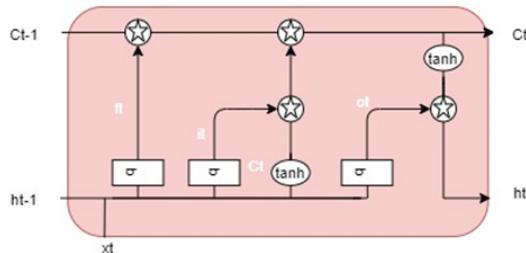


Fig. 7. General LSTM Model

The labels  $Ct-1$ ,  $ht-1$  and  $xt$  represent the input values for the LSTM diagram shown in Fig. 7. The stars represent pointwise operations, and the rectangular boxes are neural networks. Each arrow represents a vector movement.  $Ct-1$  to  $Ct$  represents the cell state and is the key to any LSTM model. Each pointwise operation is followed by a quadrate which were seen by the sigmoid function. Together the pair is called a gate, and their prime purpose is to allow data to flow through. Each LSTM model consisted of three gates to control the flow of information. The outputs of these gates were either 0 or 1. Which represent to let information through or not, respectively. Information flowed from  $ht-1$  and  $xt$  through the gates one and two, followed by the hyperbolic tangent operator. Then finally through the third gate and the second hyperbolic tangent operator followed by the output [31].

Our LSTM model has two layers, both of which uses the ReLU activation function, and a kernel optimizer. The model also has a loss and momentum function to prevent over fitting and help with model performance.

#### H. Performance Metrics

In the experiment that was conducted a total of four metrics were used. The descriptions of the evaluation metrics are [8]:

- *Precision:* Precision is the number of most relevant instances amongst those that were retrieved. It is a metric used in our findings to determine how correct our model is and how many true predictions were made.
- *Recall:* Recall measures the actual number of relevant instances made, without taking those that were retrieved into accord [8]. This metric determines the number of positive predictions made by the model.
- *F1-score:* Also known as the F score, these metrics are used in binary classifiers. They determine how accurate each model is according to the data-set.
- *Model accuracy:* This is the output accuracy taken from the model training. It determines how accurately the model has been trained.

#### IV. RESULTS AND DISCUSSION

In this section, we present the results of the comparative study that investigated emotion recognition using deep learning algorithms. Five deep learning models were considered (i.e. CNN, MLP, GAN, RNN and LSTM) and compared in terms of four metrics (accuracy, F1-score, precision and recall). The learning rates of each model produced was also discussed. Details on the configuration of each algorithm studied was provided in Section III-D.

The results of the study are summarised in Table II, which gives results broken down by model and data set used. Analysing the results presented in Table II, the following points are noted:

- The LSTM model fails to learn for both data-sets and is not suitable for performing FER. This is intuitively understandable, as the LSTM model is deep learning algorithm that is designed primarily for sequential or chronological data [29]. As such, poor performance was understandable when considering independent, unrelated images such as those used in this study.
- The CNN, MLP and GAN models are overfitted to the respective data-sets and do not produce trained models that can be applied to generalised data. This makes these models unsuitable for performing facial emotion recognition.
- The RNN learns appropriately and achieves a 72% accuracy on the larger of the two data-sets (Data-set 2). The accuracy and loss curves indicate that this accuracy can be improved by further training, but this was not done in this study due to the computational intensiveness of training the model and its low learning rate. The RNN performs best in terms of precision, recall and F1-score.

#### V. CONCLUSION

In this study, we considered five deep learning algorithms for facial emotion recognition, with the overall objective of utilising deep learning to improve marketing business functions by soliciting more accurate feedback on CX. The algorithms studied were the CNN, MLP, GAN, RNN and LSTM. Although literature often uses the CNN for facial emotion recognition studies, in our study the CNN overfits to training data and the RNN is found to be more suitable. The designed RNN is computationally intensive to train, and training in this study was terminated prematurely. The model achieved a 72% testing accuracy on a 21 000-image subset of the FER 2013 data-set, and indicators are that more training could improve this accuracy. This motivates for more intense studies that design RNN-based computer vision systems for facial emotion recognition. We recommend that future works in this area use cloud computing technologies to overcome the limitations of computational intensiveness. In so doing, a larger and more varied data-set could be evaluated which may influence the performance of the models evaluated. It would also be interesting to consider more emotion classes in future studies to cover a broader spectrum of human reactions.

TABLE II. SUMMARY OF RESULTS

Model	Data-set	Accuracy	Precision	F1-score	Recall	Learning and Fit
CNN	1	0.68	0.73	0.68	0.68	Loss-epoch curve shows evidence of severe overfitting.
	2	0.70	0.70	0.70	0.70	Loss-epoch curve shows evidence of severe overfitting.
MLP	1	0.56	0.66	0.57	0.57	Loss-epoch curve shows evidence of moderate overfitting.
	2	0.57	0.63	0.56	0.56	Loss-epoch curve shows evidence of severe overfitting.
GAN	1	0.68	0.67	0.60	0.60	Loss-epoch curve shows evidence of moderate overfitting.
	2	0.65	0.72	0.66	0.66	Loss-epoch curve shows evidence of severe overfitting.
RNN	1	0.60	0.78	0.71	0.71	Learning rate is slow, further training could improve accuracy.
	2	0.72	0.70	0.73	0.73	Learning rate is slow, further training could improve accuracy.
LSTM	1	0.33	0.50	0.33	0.33	Model does not learn.
	2	0.35	0.30	0.46	0.46	Model does not learn.

## FUNDING

All funding to support this research was provided by the Durban University of Technology.

## DATA AVAILABILITY STATEMENT

The data on facial emotion recognition 2013 (FER-2013) is available online at <https://www.kaggle.com/ananthu017/emotion-detection-fre>

## ACKNOWLEDGMENT

The authors would like to acknowledge Lerissa Gounden (BEd) of the University of Johannesburg for editing and Ghulam Masudh Mohammed (BICTH) of the Durban University of Technology for their respective inputs.

## CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

## REFERENCES

- [1] H. Zaim, M. Ramdani, and A. Haddi, “E-crm success factors as determinants of customer satisfaction rate in retail website,” *International Journal of Computer Information Systems and Industrial Management Application*, vol. 12, pp. 082–092, 2020.
- [2] S. Harvey, “The power of emotional marketing: Once more with feeling,” <https://fabrikbrands.com/the-power-of-emotional-marketing/>, n.d., accessed: 2022-04-02.
- [3] S. Ceccacci, A. Generosi, L. Giraldi, and M. Mengoni, “An emotion recognition system for monitoring shopping experience,” in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 2018, pp. 102–103.
- [4] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Emotion recognition using facial expressions,” *Procedia Computer Science*, vol. 108, pp. 1175–1184, 2017.
- [5] W. Yi, Y. Sun, and S. He, “Data augmentation using conditional gans for facial emotion recognition,” in *2018 Progress in Electromagnetics Research Symposium (PIERS-Toyama)*. IEEE, 2018, pp. 710–714.
- [6] K. Siau and Y. Yang, “Impact of artificial intelligence, robotics, and machine learning on sales and marketing,” in *Twelve Annual Midwest Association for Information Systems Conference (MWAIS 2017)*, vol. 48, 2017, pp. 18–19.
- [7] H. Liang, X. Sun, Y. Sun, and Y. Gao, “Text feature extraction based on deep learning: a review,” *EURASIP journal on wireless communications and networking*, vol. 2017, no. 1, pp. 1–12, 2017.
- [8] G. M. Mohamed, S. S. Patel, and N. Naicker, “Data Augmentation for Deep Learning Algorithms that perform Driver Drowsiness Detection,” *Scientific Programming*, 2022, under review.
- [9] P. Ładyżyński, K. Zbikowski, and P. Gawrysiak, “Direct marketing campaigns in retail banking with the use of deep learning and random forests,” *Expert Systems with Applications*, vol. 134, pp. 28–35, 2019.
- [10] D. Koehn, S. Lessmann, and M. Schaal, “Predicting online shopping behaviour from clickstream data using deep learning,” *Expert Systems with Applications*, vol. 150, p. 113342, 2020.
- [11] L.-C. Cheng and S.-L. Tsai, “Deep learning for automated sentiment analysis of social media,” in *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, 2019, pp. 1001–1004.
- [12] B. C. Ko, “A brief review of facial emotion recognition based on visual information,” *sensors*, vol. 18, no. 2, p. 401, 2018.
- [13] V. Tümen, Ö. F. Söylemez, and B. Ergen, “Facial emotion recognition on a dataset using convolutional neural network,” in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE, 2017, pp. 1–5.
- [14] Y. Luo, L.-Z. Zhu, and B.-L. Lu, “A gan-based data augmentation method for multimodal emotion recognition,” in *International Symposium on Neural Networks*. Springer, 2019, pp. 141–150.
- [15] A. Sepas-Moghaddam, A. Etemad, F. Pereira, and P. L. Correia, “Facial emotion recognition using light field images with deep attention-based bidirectional lstm,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 3367–3371.
- [16] J.-F. Reyes, “Effect of emotion on marketing landing page conversion,” Ph.D. dissertation, University of Baltimore, 2016.
- [17] G. Yolcu, I. Oztez, S. Kazan, C. Oz, and F. Bunyak, “Deep learning-based face analysis system for monitoring customer interest,” *Journal of ambient intelligence and humanized computing*, vol. 11, no. 1, pp. 237–248, 2020.
- [18] H. Zhang, A. Jolfaei, and M. Alazab, “A face emotion recognition method using convolutional neural network and image edge computing,” *IEEE Access*, vol. 7, pp. 159 081–159 089, 2019.
- [19] X. Zheng, M. Wang, and J. Ordieres-Meré, “Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0,” *Sensors*, vol. 18, no. 7, p. 2146, 2018.
- [20] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Applied Soft Computing*, vol. 97, p. 105524, 2020.
- [21] M. M. T. Zadeh, M. Imani, and B. Majidi, “Fast facial emotion recognition using convolutional neural networks and gabor filters,” in *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*. IEEE, 2019, pp. 577–581.
- [22] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, “Hybrid deep neural networks for face emotion recognition,” *Pattern Recognition Letters*, vol. 115, pp. 101–106, 2018.
- [23] P. Singh, “Visual-Recognition-using-CNN,” <https://github.com/pradeepsinng/Visual-Recognition-using-CNN/>, 2018, accessed: 2022-04-09.
- [24] D. Han, Q. Liu, and W. Fan, “A new image classification method using cnn transfer learning and web data augmentation,” *Expert Systems with Applications*, vol. 95, pp. 43–56, 2018.
- [25] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–18, 2019.
- [26] Z. Peng, J. Li, and Z. Sun, “Emotion recognition using generative adversarial networks,” in *2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC)*. IEEE, 2020, pp. 77–80.

- [27] O. Krestinskaya, B. Choubey, and A. James, "Memristive gan in analog," *Scientific reports*, vol. 10, no. 1, pp. 1–14, 2020.
- [28] S. Shamsher, "Digit-recognition-using-GAN," <https://github.com/SaraShamsher/Digit-recognition-using-GAN>, 2020, accessed: 2022-04-09.
- [29] N. K. Manaswi, N. K. Manaswi, and S. John, *Deep learning with applications using python*. Springer, 2018.
- [30] J. H. Bappy, J. R. Barr, N. Srinivasan, and A. K. Roy-Chowdhury, "Real estate image classification," in *2017 ieee winter conference on applications of computer vision (wacv)*. IEEE, 2017, pp. 373–381.
- [31] A. Sinha, "Understanding of LSTM Networks," <https://www.geeksforgeeks.org/understanding-of-lstm-networks/>, 2021, accessed: 2022-04-02.