School of Computer Science and Mathematics

6100COMP Project

Final Report Submitted by

Morgan Powell

ID Number : 890805

Computer Science

Title

Research on Facial Expression Recognition(FER) techniques and Picking the best Deep Learning Algorithm to inject into Android App.

Supervised by

Sud, Sudirman

Submitted on

5th of May 2022

Abstract

Facial expression of emotions are considered as one of the key factors in human social interactions. Recent studies have specifically shown that facial expressions are largely used as tools to specifically understand social interactions rather than personal emotions. Thus, the facial expressions credibility assessment, specifically, the distinction of posed expressions (deceptive/ volitional /deliberate) from genuine (spontaneous) ones, is a fundamental yet challenging part of facial expression recognition. With the present advancement in computer vision and machine learning techniques, substantial developments have been made in automatic detection of posed and genuine facial expressions in recent years.

As such, the works presented in this paper sets out to develop a reliable facial recognition system by testing a number of algorithms for facial expression recognition and picking the best one to be implemented in a phone application. To achieve that, the paper first analyses the relevant knowledge and systems, including a number of SVP (spontaneous versus posed) facial expression databases and a range of machine learning algorithms and techniques, neural networks nodes, and computer vision-based detection techniques, among others. Additionally, the paper also discusses the various factors that are capable of influencing the performance of the detection algorithms as well as technical challenges and open issues in this nascent discipline.

Acknowledgement

First of all, I am grateful to my supervisor Sud Sudirman, Liverpool John Moores University and the Department of Computer Science who provided me with the opportunity to undertake this remarkable and state of art project on the topic of Facial Expression Recognition. They have been a great help to me while I have been conducting a range of research, of which has opened my mind to a whole new world of knowledge.

Secondly, I am also greatly thankful to my parents, colleagues and friends who have also played a key role in the completion of this project due to their terms of encouragement and offering support. Lastly, I am also grateful to the online-based scholars and databases which were helpful in formation, implementation, and conclusion of the project.

Contents

List of figures

Chapter 1: Introduction

In its simplest form, facial expression recognition can be defined as the technology adopted to analyse sentiments by various sources, like videos and pictures, among others. According to Namba *et al*. (2021), the discipline is a part of technologies commonly dubbed ‘affective computing’, an interdisciplinary area of research concerned with the capabilities of computer to detect and interpret human emotions and affective statuses and is largely based on AI (Artificial Intelligence) technologies.

Extensively, Brown (2020) considers facial expression recognition to be a set of processes carried out by computers or humans that includes:

- Identifying a face from a scene (for instance, within an image; face detections).
- Obtaining facial aspects from the regions of the identified face (for example, identifying the structure of facial elements or defining the skin texture in a facial region; facial feature extraction),
- Analysing the motion in the facial aspects and/or any change within the facial features appearance and grouping the data into specific facial expression-interpretive classes like facial muscle instigations such as frown or smile, emotions (affect) classifications such as anger or happiness, attitude groupings such as ambivalence or like (dis)liking, among others (this is also known as facial expression interpretation).

A facial expression, on the other hand, is simply a form of non-verbal communication, which can be used for understanding human emotions. For years, deciphering such emotion expressions has been a field of interest for research by a number of fields, including psychology, Human Computer Interaction (HCI), and even biology, among others (Kouda, 2018). Of late, the high dissemination of cameras and the technological advancement in biometrics analyses and machine learning and pattern recognition have had a significant impact in the development of the facial expression recognition technology. Most corporations today, including tech giants like Google or NEC or minor ones, like Eyeris or Affectiva, are investing heavily in the technology, which indicate the growing importance of the field.

One of the universally held predictions in the world today is that computing is most likely to shift to the back, lacing itself into the structure of our daily lives and casting the human users into the forefront. To achieve that, next gen computing (that is, human computing, ambient intelligence, and pervasive computing, among others) will have to come up with human-centred user interfaces which readily respond to spontaneously occurring, all-round, humans' communications. Such interfaces will entail the capability of perceiving and understanding emotions and intentions as communicated by affective and social signals (Shaham *et al*., 2020).

Driven by that vision of the future, the work carried out in this project uses tensorflow to build a machine learning-based facial expression recognition application. It also includes tests, which are specifically done using 149 images with 20-75 kb (kilobytes), which means that the 50mb (megabytes) of images are tested. This is broadly aimed at facial expression recognition and automated analyses of nonverbal behaviour, and particularly of facial behaviour, based on concepts from human-computer interaction, pattern recognition, and computer vision. Therefore, the paper is mainly focused on the models/ algorithms used to recognise human emotions from identified human faces. The analysed data is communicated by the areas of the mouth and the eye into combined new images in different facial appearances related to the 6 general key facial emotions. The acquired output data can be used as inputs for machines with the capability of interacting with common abilities, in the contexts of developing socially intelligent system. The methodology includes various techniques and models used in recognition of facial expression information to study the region of the mouth, nose, and eyes as well as other sensitive parts to human's expression changes and which are specifically significant in decoding an emotional expression. Finally, the merged images are used as inputs to feed forward neural networks trained through various methods such as back propagation, among others. The analyses of used images facilitate it, acquire related data by combining the appropriate data in a single image and reducing the training set times while preserving classification rates. The experiment results shows the best algorithm (most accurate) to use in detecting emotion.

Undeniably, facial expressions are the most clear, spontaneously top ways for people to convey emotions, to stress and clarify what they say, to indicate their intentions, disagreements, and comprehension, among others; in short, they are useful in regulating interactions with the environment and others in their surroundings (Kouda, 2018). The automatic analyses of facial expressions afforded by this project, therefore, will be the core of a variety of next generation-computing tools including patient profiled personal wellness technologies, learner-adaptive tutoring systems, and affective computing technologies (affective and proactive user interfaces), among others.

The rest of the paper is structured in this manner: the following section, Section II includes a review of the pertinent literature with emotion recognition in facial expressions, emotions in machines, machine learning, and facial expression recognition using Neural-network-based techniques, followed by section III, which includes the tests and implementation of facial emotion detection models. In section IV, the test results are presented, while Section V includes a discussion of the results, recommendations, future works, and conclusion of the paper.

Chapter 2: Background research and domain analysis

Automatic emotion recognition include an extensive and crucial research area which utilises 2 distinctive subjects, that is, artificial intelligence and psychological human emotion recognition. In general, the human emotional state can be obtained from non-verbal and verbal information recorded by different sensors, e.g., from physiological signals, tone of voice, and facial changes, among others. According to Neckel & Hasenfratz (2021), 7 percent of the emotional information is verbal, 38 percent is vocal, and 55 percent is visual. Change of expressions in the face in the course of communications are some of the first signs which communicate the emotional states. This is the reason why most scholars have substantial interests on this modality. According to Hameed (2014), to ensure of a better classification, the facial features have to be extracted, which can be a sensitive and a difficult task.

The exploration into emotions and their association with facial expressions can be traced back to Ekman and Freisen (2004) who are considered the first modern scientists to have ventured in facial expression. Their work included analysing a developed Facial Action Coding System (FACS) whereby facial movements were defined by AUs (Action Units). The system broke down the human face into forty-six action units with each being coded with one or more facial muscles.

In comparison to other modalities, the automatic FER (facial expression recognition) is the most studied by scholars although it is not easy task because it includes each person presenting their emotions by their own way. Moreover, there are a number of challenges and obstacles existing in this area which, according to Neckel & Hasenfratz (2021) should not neglect things such as the individual background, gender, age, luminosity, and variations of head poses, as well as the occlusion problem as result of skin illnesses, scarf, and sunglasses, among others.

Moreover, Barbosa *et al.* (2019) explains that there are a number of traditional methods used for the facial features extraction, including texture and geometric features, e.g., Gabor wavelet, local directional patterns (LDA), FAC (facial action units), and LBP (local binary patterns), among others. Today, deep learning is considered one of the very efficient and successful approaches in facial expression recognition, particularly due to

the results gotten with its architectures that allow facial features to be automatically extracted and classified using CNN (convolutional neural network) and RNN (recurrent neural network). According to Barbosa *et al*. (2019), this is what drove scientists to begin utilising the technique for human emotions recognition. There are also some efforts which have been made by scientists on deep neural network architecture development, which have produced very satisfactory result in the field.

Available facial databases

One of the key factors of the deep learning success is the neuron network training with examples. Today, there are a number of FER (facial expression recognition) databases that are now availed to the public and can be used by researchers for accomplishing this task. Each one of these databases is unique from the others in terms of the size and number of videos and images, face poses, population, and variations of the illumination. In Table 1 below, some of these databases are presented, including their descriptions as well as the emotions they include.

Database	**Description**	**Emotions**
MultiPie	Over 750,000 images recorded by 15 views and 19 radiance levels	Surprise, scream, squint, happy, neutral, disgust, and anger
MMI	2900 videos, specify the offset, apex, onset, and neutral	neutral and 6 basic emotions
GEMEP FERA	289 images classifications	Happy, relief, sadness, fear, and anger
SFEW	700 images including varying head poses, illumination, occlusion, and ages	neutral and the 6 basic emotions

CK+	593 videos for non-posed and posed expressions	Neutral, contempt, and the 6 basic emotions
FER2013	35,887 grayscale images collected from Google image searches	Neutral and the 6 basic emotions
JAFFE	213 grayscale images modelled by ten females of Japanese heritage	Neutral and the 6 basic emotions
BU-3DFE	2500 three dimensional facial images captured on 2 views, +45° and -45°,	Neutral and the 6 basic emotions
CASME II	247 micro-expressions classifications	Regression, surprise, disgust, and happy, among others
Oulu-CASIA	2880 videos recorded in 3 distinct illumination settings	6 basic emotions
AffectNet	Over 440.000 images gathered from the internet	Neutral and the 6 basic emotions
RAFD-DB	3000 images from the real world	Neutral and the 6 basic emotions
RaFD	8040 images with varying genders, ages, and face poses	Neutral, contempt, and the 6 basic emotions, contempt and neutral

Deep learning

Deep learning (DL), which is also referred to as deep structured learning, is one of the key parts of the extensive ML approaches family (Malhotra, 2018). It is based on ANN with representation learning, which could be unsupervised, semi-supervised or supervised.

According to Malhotra (2018), deep-learning architectures like convolutional neural networks, recurrent neural networks, deep reinforcement learning, deep belief networks, and deep neural networks, among others are all extensively applied to various fields such as material inspection, climate science, medical image analysis, drugs design, bioinformatics, machine translation, natural language processing, speech recognition, computer vision, and board game platforms, whereby they have managed to produce results analogous to and in other cases better than the performance of human experts.

In deep learning (DL), the adjective "deep" entails the utilisation of various network layers. Early works, such as by Fourie (2003), indicated that linear perceptron could not be universal classifiers; however, networks with nonpolynomial activation functions and hidden layer of unlimited widths could. Chah (2019) considers deep learning as one of the recent variations that is interested with infinite amount of layers of limited sizes, which permit optimised and practical implementation of applications, while maintaining theoretic generality under moderate condition. In DL, the layers could also be diverse and diverge extensively from a biologically-based connectionist model to ensure understandability, trainability, and efficiency of the "structured" parts.

The majority of the modern deep learning models are based on ANNS, particularly CNNs (convolutional neural networks), though they can also comprise latent variables or propositional formulas structured in layers in deep generative models like the nodes in deep Boltzmann machines and deep belief networks (Chah, 2019).

In DL, individual levels learn to convert their input data into more complex and theoretical representation to some extent. In images or face recognition applications, the raw inputs might include a milieu of pixels, in particular; abstracting the pixels and encoding edges

could be done by the first representational layer; followed by composing and encoding of the edges' arrangements that can be done by the second layer; next is the encoding of eyes and nose by the third layer; and recognition of a face in the image by the fourth layer. Notably, deep learning processes are capable of learning what features to place optimally in what levels on their own. According to Sokolov (2017), that does not remove the need for manual tuning; for instance, diverse layer sizes and different layer numbers are capable of providing different levels of abstraction.

Solopchuk & Zénon (2021) also explain that the term "deep" in "deep learning" signifies the various layers where the data is converted. More precisely, DL systems include significant CAP (credit assignment path) depths. The credit assignment path refers to the sequence of transformation from inputs to outputs. According to Malhotra (2018), it defines possibly causal relationships between inputs and outputs. In feed forward neural networks, the credit assignment path depths is those of the networks, which are the amount of hidden layers added to 1 (since the output layers are parameterised too). In recurrent neural networks, whereby signals could propagate through layers multiple times, the credit assignment path depth is theoretically unlimited (Sutskever & Hinton, 2020). Although there is no universal established threshold of depth capable of dividing shallow learning from deep learning, a lot of scholars hold that deep learning comprises credit assignment path depth greater than two. Credit assignment path of depth 2 has been indicated to be a common approximator in the viewpoint that it is capable of emulating any functions. Beyond this, extra layers do not augment the network function's approximator abilities. Deep models (credit assignment path > 2) have the capability of extracting better features as compared to shallow models and thus, extra layers can be used to effectively learn the features.

According to Sutskever & Hinton (2020), a DL architecture can be developed by use of greedy layer-by-layer methods. Deep learning facilitates the disentangling of the abstractions and picking the features enhance performance.

For supervised learning, DL techniques get rid of features engineering, by converting the information to a solid intermediary representation similar to main elements, and obtain a layered structure, which get rid of redundancies in representations.

Fong & Hong (2021) explain that DL algorithms can also be utilised in unsupervised learning tasks, which is a key advantage since unlabeled data are richer in comparison to labelled data. A good example of deep structure which can be trained with unsupervised learning is deep belief network.

Solopchuk & Zénon (2021) state that ANNs (Artificial neural networks) were inspired by distributed communication nodes and information processing in biological systems. Artificial Neural Networks, however, differ in a number of ways from biological brains. For instance, ANNs are symbolic and static while the most living organisms' biological brains are analogue and dynamic (plastic).

Neural networks

Artificial neural networks

According to Solopchuk & Zénon (2021), artificial neural networks (ANNs), also referred to as connectionist systems, refer to computing systems which borrow from the biological neural networks which comprise animal brains. These systems learn (increasingly improving their abilities) to perform tasks by studying examples, basically with no task-exclusive programming. E.g., in image recognition, they have the capability of learning to detect pictures which include dogs by analysing sample imageries which are labelled manually as "dog" or "no dog" and with the analytic result to recognise dogs in another image. They are commonly used in applications that are hard expressing using the traditional computer algorithms based on rule-based programming.

Artificial neural network are based on sets of related components referred to as artificial neurons, (similar to a biological brain's neurons). Each of the connections (synapses) between the neurons is capable of transmitting signals to other neurons (Solopchuk & Zénon, 2021). The receiving (postsynaptic) neurons have the capability of processing the signal(s) and then signalling the connected downstream neurons. Neurons might include states, basically characterised by real numbers, normally between 1 and 0. Synapses and neurons could also include weights which vary as learning continues, which is capable of increasing or decreasing the signal's strength sent downstream.

Usually, neurons are organised into various layers, capable of performing various types of conversions on their input. A signal travels to the last (outputs) from the first (inputs) layer, probably once they have gone through the layers in various instances.

The initial aim of the neural network method, as put by Mark (2017), was to find solutions to problems how human brains did. However, with time, focus was shifted to match certain mental capabilities, which led to deviations from biological processes like backpropagation, or communicating in the contrary directions and altering the networks to return this information.

Sáez Trigueros *et al*. (2021) note that neural networks are applied on a range of tasks, such as medical diagnosis, video games, playing boards, social network filtering, machine translation, and computer vision, among others. As of 2022, neural networks basically include millions of units and connections. In spite of the number being lesser than the human brain neurons, the networks are capable of performing a range of tasks at levels beyond those of human beings (such as in playing “Go” and recognizing faces, among others).

Deep neural networks

These are basically artificial neural networks (ANNs) with more than one layers between the output and the input layers. Although there are various kinds of neural networks, as put by Sáez Trigueros *et al.* (2021), they are always made up of the same elements: functions, biases, weights, synapses, and neurons, which function as the human brain and are capable of being trained as any other machine learning algorithms.

For instance, a deep neural network which is trained for recognising the breeds of cats examines the given images and calculates the probabilities that the cats in the images is of particular breeds. The users have the freedom of reviewing the results and selecting which probability the networks should show (above particular thresholds, among others) and return the anticipated labels. All mathematical manipulations thus, are considered as layers, and complex deep neural networks include a lot of layers, thus the label "deep" network (Mark, 2017).

Moreover, deep neural networks are capable of modelling complex non-linear relations. Deep neural network architectures produce compositional models whereby the objects are conveyed as layered structure of primitives. The additional layers facilitate features composition from the lower layer, possibly modelling complex data using less units in comparison to relatedly performing shallow networks. For example, research has shown that sparse multivariate polynomials are basically simpler for approximating with deep neural networks as compared to shallow networks (Sáez Trigueros *et al*., 2021).

Deep architectures comprise a lot of variants of a small number of general methods. Although the different architectures have been successful in certain domains, comparing the various architecture performances is not always possible lest they are assessed on similar data sets.

According to Zhu & Zhao (2021), deep neural networks are basically feedforward networks whereby data flows to the output layers from the input layers without having to loop back. At first, the deep neural network generates a chart of virtual neurons and allocates "weights" or random numerical values, to relationships between them. The inputs and weights are proliferated and return outputs between 1 and 0. If certain patterns were not recognised accurately by the networks, algorithms tweak the weights. This way, some parameters are made more influential by the algorithms, until they determine the right mathematical manipulations for fully processing the data.

A RNN (Recurrent neural network), whereby data is capable of flowing in any of the directions, is utilised in applications like language modelling. They are particularly useful and effective for long short-term memory.

A CNN (Convolutional deep neural network) can be applied in a wide range of fields, including computer vision and acoustic modelling for ASR (automatic speech recognition) (Laddha & Kumar, 2022).

Issues

Similar to artificial neural networks, a lot of challenges can emerge as a result of naively trained deep neural networks. The 2 common challenges are computation time and overfitting (Laddha & Kumar, 2022).

Deep neural networks include overfitting due to the extra abstraction layers that allow the modelling of rare dependencies in the training data. Regularisation approaches like sparsity (l_1-regularization), weight decay (l_2-regularization), or Ivakhnenko's unit pruning can be applied in the course of training to deal with overfitting. Alternately, dropout regularisation gets rid of units from the hidden layers at random during training to help in exclusion of rare dependencies. Lastly, data can be enhanced using techniques like rotating and cropping such that the small training sets can be enlarged in size to decrease the risks of overfitting (Fong & Hong, 2021).

Deep neural networks have to consider a lot of training parameters, like the initial weights, the learning rate, and the size (amount of units per layer and the amount of layers). Brushing through the parameter spaces for optimum parameters might not be practicable because of the cost in computational resources and time. A number of tricks, including batching (computation of the gradients on various training examples simultaneously instead of individual sample) can be used for speeding up computations. Large processing abilities of multiple-core architectures (like Intel Xeon Phi or GPUs) have generated substantial speedup in training, due to the aptness of these processing architectures for the vector and matrix computations.

Alternatively, developers could find other kinds of neural networks that include more convergent and uncomplicated training algorithms. For instance, cerebellar model articulation controller (CMAC) is one example of these neural networks. It does not necessitate randomised initial weights or learning rates for cerebellar model articulation controller. The training processes converge in a single step with a new data batch, and the training algorithm computational complexity is linear with regard to the involved amount of neurons (Su & Qi, 2018).

All the issues mentioned above are a very few of them that a developer face during the solution of this problem. Different type of light conditions and the resolution of the camera is also an issue in this domain. Different age group and different demography of the person are are main constraint that we cannot make a universal model for all of the population on the earth.

Applications

According to Najafian & Russell (2020), large-scale automatic speech recognition is the initial and most common effective cases of deep learning. Long Short-Term Memory recurrent neural networks could learn "Very Deep Learning" tasks which include multi-second intervals including speech episodes divided by multiple separate time steps, whereby individual time steps correspond to roughly 10 ms. Long short-term memory with forget gates is good with conventional speech recognisers on particular tasks.

The initial speech recognition success was based on small-scale TIMIT-based recognition tasks. The datasets contain six hundred and thirty speakers from 8 key American English dialects, in which individual speakers read ten sentences. The small size is helpful in allowing attempts of varied configurations. More prominently, the TIMIT tasks involve phone-sequence recognitions, which, contrary to word-sequence recognitions, allow weaker receiver bigram language models. That allows the strengths of the acoustic modelling features of speech recognition to be more simply analysed.

The late 1990s' introduction of deep neural networks for speaker recognition, the 2009 - 2011 speech recognition, and the 2003 – 2007of long short-term memory facilitated development in 8 key areas:

- Scale-out/up and enhanced deep neural networks decoding and training
- Processing of features by deep models including strong understanding of the underlying procedures.
- Adjustment of deep neural networks and associated deep models.
- Transfer learning and multi-task by deep neural networks and associated deep models
- Convolutional neural networks and the way they can be designed to effectively utilise field understanding of speech.
- Recurrent neural networks and their rich long short-term memory variations.
- Other forms of deep models as well as tensor-based models and incorporated deep discriminative/ generative models.
- Sequence discriminative training.

Every top commercial speech recognition system (including, the various Nuance speech products, iFlyTek voice search, Baidu, Apple Siri, Google Now, Amazon Alexa, Skype Translator, Xbox, and Microsoft Cortana, among others) are built on deep learning.

Image recognition

One of the commonly used evaluation sets for images classification is the Modified National Institute of Standards and Technology (MNIST) database data sets. The database includes handwritten numbers and over ten thousand test examples and sixty thousand examples. Like TIMIT, MNIST's manageable smaller size allows for multiple configuration tests by users (Wen, 2019).

Deep learning-based image recognitions are considered as superior in terms of generating more accurate result in comparison to humans. According to AL-Oudat *et al*. (2022), this has been proven twice, with the first confirmation being on 2011 with traffic signs recognition and with human faces recognition in 2014. Moreover, deep learning-trained cars are now capable of interpreting 360° camera angles. The FDNA (Facial Dysmorphology Novel Analysis), which is used for analysing instances of human deformity associated with large genetic syndromes databases is another good example.

Visual art processing

Closely associated with the advancements which have been experienced in image recognition is the growing applications of DL methods to a number of tasks in visual art (Koroscik, 2022). In this, deep neural networks have proven their capability in, for instance:

- Recognising the style periods of various paintings
- Neural Style Transfers – determining the styles of particular artworks and applying them in visually satisfying manners to random videos or photographs.
- Producing outstanding imageries based on arbitrary visual input fields.

Natural language processing

According to AL-Oudat *et al.* (2022), since the early 2000s, neural networks have been utilised for the implementation of language models. Then, long short-term memory facilitated the improvement of language modelling and machine translation.

Other major approaches in that field are word embedding and negative sampling (Mochihashi, 2020). Word embedding, like word2vec, can be considered as representational layers in a DL architectures which transform atomic words into positional representations of the words relative to others in the datasets; the positions are characterised as points in vector spaces. The use of word embedding as recurrent neural network input layers allow the parsing of phrases and sentences by the networks using efficient compositional vector grammars. Compositional vector grammars can be considered as PCFGs (probabilistic context free grammars) implemented by recurrent neural networks. Recursive auto-encoders based on word embedding have the capability of detecting paraphrasing and assessing similarities in sentences (Dyrka *et al*., 2019). Deep neural architectures deliver the top results for text classification, writing style recognition, contextual entity linking, machine translation, spoken language understanding, information retrieval, sentiment analysis, and constituency parsing, among others.

Recent advancements are capable of generalising words embedding to sentences embedding.

GT, popularly known as Google Translate, utilises significant end-to-end LSTM networks, GNMT (Google Neural Machine Translation), which utilises example-based machine translation methods whereby the systems learn from thousands of related instances. It functions basically by translating full sentences at a time, instead of fragments. Moreover, GT supports more than 100 languages. GNMT encodes the sentence semantics instead of basically memorising individual translations of different phrases. English is used as an intermediary between most language sets in Google Translate (Dyrka *et al*., 2019).

Toxicology and drug discovery

According to Rupa (2021), majority of the candidate drugs end up failing to earn regulatory approvals. Such failures can be caused by a number of things, including unanticipated toxic effect, undesired relations (off-target effect), or inadequate efficiency (on-target effects). As such, researchers have lately been using deep learning for predicting the biomolecular off-targets, targets, and impacts of toxic environmental chemicals in drugs, household products, and nutrients.

An example of DL application in this context is AtomNet, which is a DL structure-based rational drug design system utilised for predicting new candidate disease targets biomolecules like multiple sclerosis and the Ebola virus (Rupa, 2021).

The first graph neural networks for predicting different molecule properties in a big toxicology data set were used in 2017. Two years later, in 2019, generative neural networks were utilised for generation of molecules which were experimentally confirmed by use of mice (Rupa, 2021).

Customer relationship management

Deep reinforcement learning is utilised for approximation of the values of likely direct marketing practices, defined in RFM (Recency, Frequency, Monetary Value) variables terms. The projected values function has revealed to include natural interpretations as client lifetime values (AL-Oudat *et al.*, 2022).

Recommendation systems

Recommendation systems use DL for the extraction of meaningful aspects for underlying factor models for content-based journal and music recommendations. Multi-view DL is also used for learning the preferences of the users from different areas. The models use hybrid content-based and collaborative approaches and enhance recommendation in various tasks (Schedl, 2019).

Bioinformatics

Auto encoder artificial neural networks have been applied in bioinformatics, for predicting gene-function relations and gene ontology annotations (Pevzner, 2019).

DL is also applied in medical informatics for predicting the quality of sleep according to data from wearable technologies and health complication predictions based on electronic health data records.

Medical image analysis

DL has proved reliable in producing viable results in medical applications like image enhancement, organ segmentation, lesion detection, and cancer cell classification, among others (Pevzner, 2019). Modern DL techniques are proven to be highly accurate

in detection of a range of illnesses and helpful to specialists in terms of improving the diagnoses efficiency.

Mobile advertisements

Getting the right audience for mobile advertisement can be challenging as explained by Koroscik (2022). This is because a lot of data points have to be taken into consideration and analysed before the creation and utilisation of a target group in advert serving by the advert servers. DL has proved to be useful in interpreting big, multi-dimensioned advertisement datasets. A lot of data points are gathered in the course of the click-serve-request internet advertising cycles, which can produce the base of machine learning for improving advert selection.

Image restoration

DL techniques have also been found useful in inversing problems like film colourisation, inpainting, super resolution, and denoising, among others (Mochihashi, 2020). Such applications consist of learning techniques like Deep Image Prior, which train on the images that require restorations and "Shrinkage Fields for Effective Image Restorations" which train on image datasets.

Detection of financial frauds

DL methods have also been effectively applied in anti-money laundering, tax evasion detection, and financial fraud detection systems and processes.

Military

An example of this is the US Department of Defence applying DL for training robots in new tasks via observations (Dyrka *et al*., 2019).

Partial differential equations

Physics-based neural networks are utilised in solving partial differential equations in both inverse and forward problems in data-driven manners. For instance, the reconstruction of fluid flow managed by the Navier-Stokes equation. The use of physics-based neural networks does not necessitate the mostly costly mesh generation which traditional CFD approaches rely on (Fadaei & Moghadam, 2017).

Image reconstructions

The reconstruction of images involves reconstructing the underlying imageries from the image-related dimensions. Research has shown the superior and reliable performance of DL techniques in comparison to analytical approaches for a range of applications, for instance, ultrasound imaging and spectral imaging (Schedl, 2019).

Key differences between machine learning and deep learning

Although deep learning (DL) is considered as a subset of machine learning (ML), they include a number of key difference. For instance, DL differs from traditional machine learning on the kind of data that it can work with and the approaches used for learning (Kaur, 2022).

Also, ML algorithms utilise structured, labelled data for making predictions, which means that certain features are identified from the model input data and organised into tables. That does not essentially mean that they do not utilise unstructured data; it only implies that if they do, they basically follow some pre-processing for organising them into structured formats.

DL gets rid of some of data pre-processing which is basically included in ML. The algorithms are capable of ingesting and processing unstructured data, such as images and text, and they automate extraction of features, getting rid of some of the reliance on human users. For instance, assuming that there was a set of photos of various animals, and the requirement was to categorise them by “domesticated”, “wild”, et cetera, DL algorithms can be useful in determining what features (such as teeth) are key for distinguishing the different animals from one another. In ML, that order of features is manually determined by human experts (Kaur, 2022).

Besides, through the backpropagation and gradient descent processes, the DL algorithms adjust and fit themselves for accuracy, which allow them to come up with predictions regarding a new image of an animal with enhanced accuracy.

ML and DL models include various kinds of learning too, which are normally categorised as reinforcement learning, unsupervised learning, and supervised learning (Sharma,

2020). Supervised learning utilises labelled datasets for categorising or making predictions, which require some form of human intervention for correctly labelling input data. Unsupervised learning, on the other hand, do not necessitate labelled datasets, and rather, it recognises data patterns, grouping them by any unique features. Lastly, reinforcement learning includes models learning to be more accurate to perform actions in environments according to feedbacks so as to maximise the rewards (Sharma, 2020).

Facial emotion recognition using deep learning

Even with the noteworthy effectiveness of conventional facial recognition approaches in extracting handwritten features, researchers have started directing their attention to deep learning approaches particularly because of their high accuracy in automatic facial recognition. In that context, the following paragraphs in this section include some of the proposed DL methods for obtaining superior detection. The methods have been trained and tested on various sequential or static databases.

Mollahosseini *et al.* (2016) explains a deep convolutional neural network for facial expression recognition across various existing databases. After extraction of the facial features from the data, the images shrunk to 48x48px and enhanced it with deep learning techniques. The used architecture included 2 convolution-pooling layers and 2 inception styles modules, which include CNN layers size 5x5, 3x3, and 1x1. This enhanced its performance thanks to the convolution layers locally applied. Moreover, the technique also made it possible reducing the problem of over-fitting.

In their work in facial expression recognition, Chen et al. (2020) explain the impacts that data pre-processing before the network is trained it could have in provision of emotion classification with superior performance. To explain that, CNN, including 2 convolution pooling layers with 2 fully connected 7 and 256 neurons were used. The process involved a number of steps, including intensity normalisation, down sampling with 32x32px, cropping, rotation correction, and data augmentation. The best acquired weights at the training stages were utilised at the test phase. Its capability was assessed in 3 available databases: BU-3DFE, JAFFE, and CK+. The study also indicates how bringing together

all of those pre-processing stages could be more efficient as compared to applying them individually. The pre-processing methods were also employed by (Mollahosseini et al., 2016) in their proposal of an innovative CNN that can be used to detect the AUs of the faces. The networks utilise 2 convolution layers, which are followed individually by a max pooling and end with 2 wholly connected layers which show the number of activated AUs.

To deal with the problem of disappearance or explosion gradient, Dyrka et al., (2019) proposed a novel architecture convolution neural network using Sparse Batch normalisation, SBP, which uses 2 consecutive convolution layers at the start, followed by SBP after max pooling. To get rid or alleviate the problem of over-fitting, the authors applied of 3 fully connected CNNs in the middle.

In the works presented above, the researchers cited and classified the following as the basic emotions: neutral, sadness, fear, anger, surprise, disgust, and happiness, as presented in the figure below:

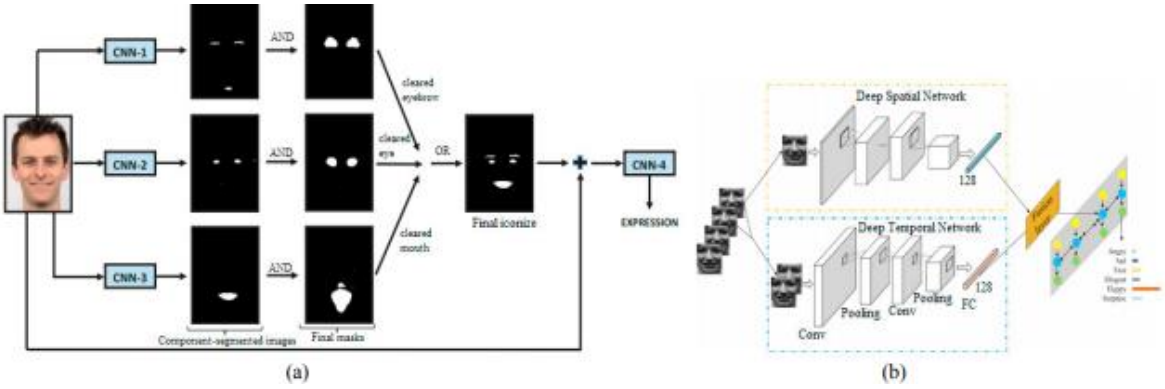


Figure 1: the various basic deep learning methods

Emotion in machines

A few years back, the concept of emotions integration in a computational system was proposed. Picard *et al.* (2001), came up with the term “affective computing” to refer to a case in which computing relate to, arise from, or intentionally influence emotions or any other affective phenomenon. One of the primary objectives of affective computing is recognizing and generating synthetic emotions that affect artificial agents. With respect to the topic under study, there is a range of key works that are focused on the emotional

interrelations between machines and humans, particularly the simulation and interpretation of emotions.

On the topic of affective computing, Andre (2021) states that research is combined to create machines capable of recognising, modelling and communicating emotions for the enhancement of human computer interactions and related work in amazing ways. For instance, Decision Affect Theory offered empirical research on the effects of expectations on generation of emotions. Extensive models like TAME produce demand for study that connect various principles and their effect on emotion and decisions making, as presented in this case. For other examples, see appendix 1, including AiSoy, Decision Field Theory, EMA, and FLAME in the appendix section.

One of 2 key field of the affective computing includes detecting emotion expression, which is also one of the issues tackled in this paper. Affective computing attempts capturing the communication of the sign that is associated with the emotion expressions, and the corresponding interpretations. The user and environment data could be recorded using a group of sensors. Once the knowledge is obtained, it is categorised in order to choose only the appropriate data for the analyses. Some of these forms of recognition are categorised in: face recognition (object of our study), natural language processing, and voice recognition, among others.

Emotion recognition in facial expressions

According to Maroti et al, (2021), modern automatic face recognition can be traced back to Suwa *et al.* (2008) with their attempts on automatic analyses of facial appearances by tracing the action areas on a sequence of images. Latest works in development of intelligent machines try identically replicating humans. However, as put by Maroti *et al*. (2021) something is still missing in the interactions between human beings and machines; “The emotional factor” is still very limited. Machines still remain these cold components which preclude a clear recognition of the emotional states of humans. In fact, the capability of displaying emotion on human-like faces is both a vital and essential phase in creating a machine that is meant for the general public access; these face are the windows in which the emotions are exhibited. Mellers *et al*. (2017) explains that there is

some correlations between every emotion that is felt and articulated in a free manner, for instance, the expression of the face is an evident exhibition of the intensities of involuntarily expressed emotions, without preceding planning or intent.

Namba *et al*. (2021) states that the key of the research in this area is exploring new methods of human-machine interactions by making computers more aware of the attentional and emotional expressions of human users. In this journey, a number of techniques for recognising emotions from faces have been presented, in harmony with that view, this set of reviews also analyses facial emotional expressions, and different physical zones based on the face muscles' mechanical movements. A more detailed review in relation to the breakdown of facial expression is presented in appendix 3.

Facial emotions recognition using neural-network-based techniques

Extracting emotions from static images facilitates the detection of various physical features like skin colour, eyebrows size, wrinkles on the eyes and the forehead, among others. Here, the neural networks are precise for the acquirement of nonlinear mapping among various data set; such analyses allow decoding of the relationships between the faces' physical aspects and their impressions. The ability of the techniques based on neural networks is the presentation of classification of facial expressions into crucial emotion classifications. The kind of assortment in 6 key emotions via techniques based on neural network was presented by Filippini *et al*. (2021), whereby, the components of the inputs to the artificial neural network relate to the distribution information of brightness obtained from a still image input. The standard percentage of recognition was 85 percent in a collection of ninety tested imageries.

Additionally, using a hybrid approach suggested by Filippini *et al*. (2021), the ANN can be reinforced by combining with HMMs (Hidden Markov Models) used in facial emotions recognition. The study includes an artificial neural network for estimating the subsequent for the discriminant Hidden Markov Models, and got positive result on the emotion classification in the lower and upper parts of the static images distinctly. The Filko & Martinović (2013) analysis utilised the analysis of main facial regions with neural networks and major component analyses, the processes were developed in a set of 15 neural

networks. Just a single artificial neural network in this set was utilised for zone detection and the remaining 14 were utilised for learning and recognising 7 basic emotions over mouth and eyes areas. The presented tests presented a 46 percent to 80 percent accurate recognition rate which was decreased to the average accuracy of 70 percent.

Successful result in emotional classifications of input still facial images were also acquired by Fadaei & Moghadam (2017) in their analysis including an illustration of outputs from 6 diverse groups of neutral emotions. In constructing the artificial neutral network, the output layer included 7 units, with each relating to a certain group of emotions; in average, the rate of recognition accuracy reached was 86%. In their work, Fadaei & Moghadam (2017) also use the neural network to perform nonlinear dimensionality reduction in the input images, since the data of interest can be found on integrated non-linear range of the high-dimensional spaces. For this stage, the algorithms come up with statistical decisions regarding the category of the detected expressions. The collection of outputs give estimations of the probabilities of the analysed expression being of the related group. The accuracy of this category increased to an average of 90.1%.

Others forms of methods like in the works of Kaur (2022) propose using Radial Basis Function Networks and Multilayer Feed-forward Neural Networks, typically utilised in pattern recognition and non-linear mapping approximation. The research utilises a group of the 7 basic forms of emotions: disgust, surprise, fear, anger, sadness, happiness, and neutral. The Euclidean is distant from the contour points in the still images and the linear coordinate from the facial feature point correspond to the dataset inputs in the neural networks. A test for the method was then done using a collection of imageries from the JAFFE database, which managed to attain an accuracy rate of 73%. Koroscik (2022) proposed a method of assessing feature areas on a fixed group of imageries, His model utilised feed-forward neural networks, whereby the inputs are the set of face features under consideration, based on the examinations found in the human facial expression study. The network model includes sets of 11 fed-forward neural networks, completely connected using vanilla neural networks. The architecture includes one hundred and five inputs per network and individual networks consist of hidden layers comprising ten nodes. Each network was trained using online back propagation. The individual networks’

outputs were merged to generate a percentage rate for the classification of the different emotions.

An experiment carried out by Shaham *et al.* (2020) indicated the capability of neuro-fuzzy network in extracting emotions in facial motions. The used approach attempts to classify intermediate and primary emotions by use of FAP (facial animation parameters) and based on the definition of the parameters. The FAPs (facial animation parameters) defined in MPEG-4 comprise an exceptionally powerful collection of parameters which facilitate an extensive range of facial motions. One of the key issues with the classification is the challenge involved in the translation of the FP (Feature Point) movements into the FAPs (Facial Animation Parameters).

Automatic Facial expressions recognition process

The machine recognition of human facial expressions primarily consists of 3 main areas (Figure. 1), namely: (a) detecting a face within the background, (b) obtaining facial expression from the identified face regions, (c) analysing the flow of facial expressions and/or the changes in appearance in the facial expressions, and categorising that data into some facial- appearance - interpretive groups (such as facial muscle behaviours and emotions, among others).

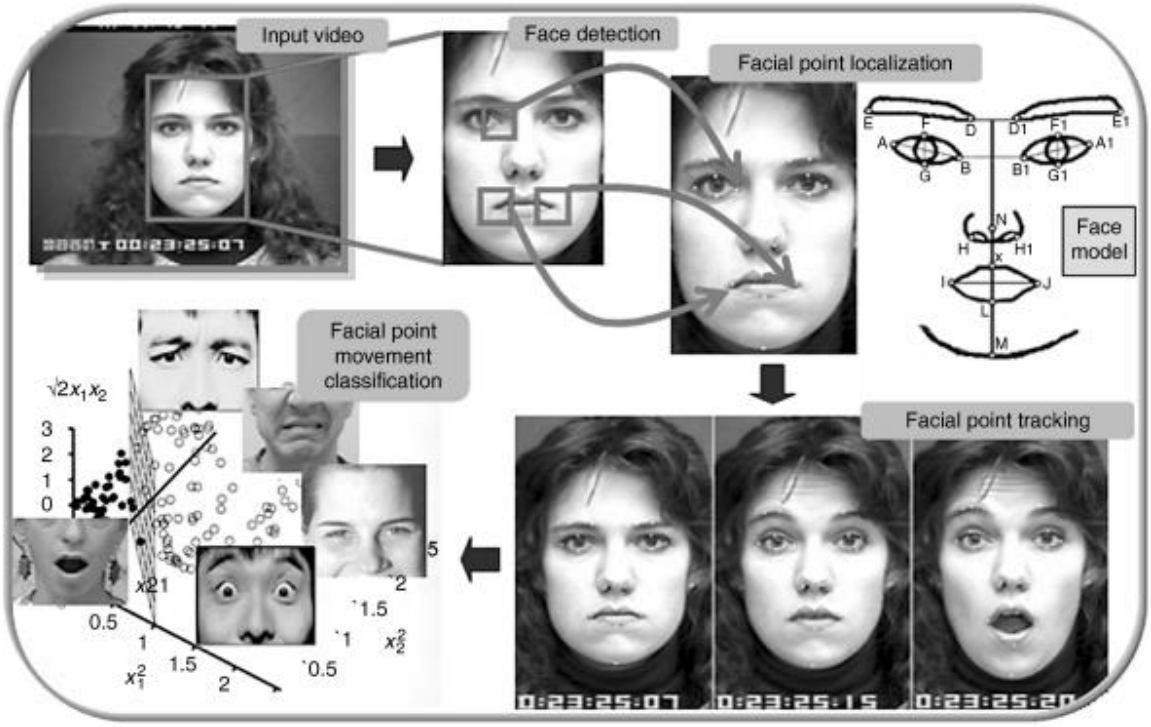


Figure 2: depiction of an automatic, geometric-feature-based system for recognition of facial expressions.

In machine vision, the problem of recognising a face can be considered as a segmentation problem while in pattern recognition, it would be seen as a detection problem. In essence, it entails identifying all areas in the scene which include a human face, including, lighting conditions, head pose variations, occlusions, and clutter, (Liu et al., 2017). Some of the involved problems in recognising and classifying the faces can be issues to do with face detection and face localization, which should be solved by associated algorithm. The existence of non-rigid movement because of facial expressions and higher levels of discrepancy in texture, colour, and facial sizes make the problem even more complex. As a response, various techniques have been created to detect faces in still images as explained by Wen (2019). But, majority of them are only capable of detecting an upright face in near frontal or frontal views. Perhaps the top universally utilised face finder in automated facial expressions analyses is the system proposed by Jones & Viola (2014) capable of face recognition in real-time.

Schedl (2019) states that the issue of feature extraction can be considered a dimensionality reduction problem (in pattern recognition and machine vision). It entails converting the input information into moderated representation group of traits that train the related data from the input information. Schedl (2019) further explains that the issue of facial feature extractions from an input image can be divvied into at least 3 aspects: (a) Are the features all-inclusive (extending over the entire face) or analytic (covering only some parts of the face)?; (b) has temporal information been utilised?; (3) Are the features volume or view-based (two dimensional/ three dimensional, etc.)?.

In view of this glossary, the majority of the proposed facial expression recognition approaches are focused on 2D, analytic, or static facial feature extraction. The commonly used obtained facial expressions are either geometric aspects like the shape of the face elements (like mouth as well as eyes, among others) and the location of facial fiducial point (the corner of the mouth and eyes, among others), or appearance aspects that represent the facial skin texture in certain facial regions such as furrows, bulges, and wrinkles, among others. Appearance-based aspects are such as learned image filters from traits based on edge-oriented histograms, integral image filters (Haar-like filters and box-filters), Gabor filters, LFA (Local Feature Analysis), PCA (Principal Component Analysis), and ICA (Independent Component Analysis), among others.

Moreover, a number of works have also been described by Rupa (2021), which utilise both appearance and geometric features. Such approaches are called hybrid methods. In these methods, however, Rupa (2021) adds that the appearance features always outperform the geometric features –based methods using, for instance, eigenfaces or Gabor wavelets. There are more cases by different researchers which show that geometric features are capable of outperforming the appearance based features. Yet, it still is considered that utilising both appearance and geometric features may be the best option with some facial expression systems.

Facial muscles contractions, which are the source of facial expressions, instigate movement of the face skin and a change in the position and/or look of facial expressions (for example, contractions of the Corrugator muscles induce frowns and cause one eyebrow to move toward the other eyebrow, typically creating wrinkles between them; as

shown in Figure 3 below). Such a change could be identified by evaluating optical flows, facial-points- or facial components-contour-tracking result, or by the use of a set of trained classifiers for making decisions on the occurrence of specific variations (such as, if the nasolabial furrows are deep or not) according to the occurred facial expressions. The optical flow approaches to defining facial motions include the benefit of not needing facial feature extraction stages of processing. Solid flow of data exists all through the whole facial region, irrespective of the availability of facial elements, as well as in the regions of smoother textures like the forehead or the cheeks. Since optical flows are the observable results of movements and are given in form of velocity, they can be utilised in directly representing the facial expressions. Majority of researchers have started adopting this approach according to Kaur (2022). Until the recent years, conventional optical flow methods were, arguably, typically used to track facial feature contours and points too. To address the drawbacks integral in optical flow approaches like changes in illumination, clutter, occlusion, sensitivity to noise, and the accumulation of errors, recent works in automatic facial expression recognition utilise sequential states estimation approaches (like Particle filter and Kalman filter) for tracking facial feature points in series of image (for example, (Mollahosseini *et al.,* 2016)).

Figure 3: Corrugator muscles contraction facial appearance

Ultimately, the tracked changes in facial components contours, facial characteristic points tracked movements, dense flow of information, and/or extracted appearance traits are

interpreted into descriptions of the shown facial expressions. The descriptions (interpretation of facial expressions) are normally presented either in form of exhibited affective state (emotion) or in form of the utilised face muscle to display the facial expressions. That is as a direct result of 2 key facial expression measurement approaches in psychological study: signs and message assessment. The assessment of the message purposely aims at inferring what is behind exhibited facial expressions, like personality or emotion, while the assessment of signs helps in describing the ‘‘surface’’ of the exhibited behaviour, like facial component shapes or facial movements. Therefore, brow frowns can be said to be ‘‘anger’’ in messages estimation and like facial movements which lower and pull the eyebrows towards each other in signs-assessment method. Although messages assessment mainly involves interpreting, sign assessment tries being more objective, which leaves inferences regarding the communicated message to top ranking decision making. In general, the typically utilised face expression descriptor in messages assessment method are the 6 main emotions (surprise, disgust, anger, happiness, sadness, and fear, as indicated in figure 4 below) according to Mochihashi (2020) and other emotion scholars, who imply that such emotions are commonly exhibited and recognised from facial expressions. The common face action descriptor used in signs assessment methods are the Action Units (AUs) presented in the Facial Action Coding System (FACS). Most of the developed facial recognition analysers to date are more focused on human facial emotion analyses and try recognising a smaller group of prototypic facial emotions expressions such as anger and happiness. In addition, there are a number of promising model systems capable of recognising Action Units in face images which are generated intentionally and even some have been making attempts toward recognition of naturally exhibited Action Units (Sharma, 2020). Although the traditional methods utilise simple techniques such as machine learning approaches and expert rules, including neural networks for classifying the related data from the input information into certain facial-expressions interpretive classes, modern (and basically more innovative) approaches utilise ensemble, statistical, and probabilistic learning methods, which are considered specifically fitting for automatic facial expression recognitions from series of face images.

Figure 4: the 6 basic emotions prototypic facial expressions (right-to-left from bottom row): surprise, fear, anger, sadness, happiness, and disgust.

Evaluation of performance of automatic facial expression recognition systems

According to Samadiani *et al.* (2019), the 2 main features of assessing performances of developed automated facial expressions recognisers are the test dataset/utilised training along with the employed test approach.

One of the prerequisites in development of robust automatic FER (facial expression recognisers) is having sufficient labelled information of the objective human facial behaviours. Malathi *et al.* (2019) show that with the right three dimensional alignment of the face, at least fifty training samples are required for adequate performance (around 80 percent accuracy) of ML approaches to recognise certain facial expressions. It is difficult recording and/or collecting natural facial behaviours since they are hard to produce, are temporary, and full of subtle context-based variations. Furthermore, manual labelling of

natural facial behaviour for more accuracy can be very expensive, error-prone, and time consuming. Because of those issues, most of the current literatures on automatic FER are built on the ‘‘artificial’’ information of intentionally exhibited facial behaviours, prompted by the subjects being asked to present a sequence of facial expressions using cameras. Some of typically utilised publicly accessible, annotated posed facial expressions datasets are such as the JAFFE database, MMI facial expression database, and Cohn-Kanade facial expression database, among others. However, Malathi *et al.* (2019) research indicates that intentional (posed) behaviours differ in timing and appearance as compared to the ones that occur in everyday life. For instance, a deliberate smile has bigger amplitudes, more short-lived durations, and faster offset and onset velocities in comparison to a spontaneously occurring smile. As such, it is not shocking that methods which are trained on intentional as well as basically exaggerated behaviour to flop in generalising the complexities of expressive behaviours existing in the real life situations. Several attempts have been made in a bid to tackle the universal absence of reference sets of visual and/or audio human spontaneous behaviour recordings by developing datasets like those. Some of the typically utilised, publicly accessible, annotated spontaneous human behaviour recording datasets are such as MMI-Part2 database, UT Dallas database, and SAL dataset.

Another common evaluation in machine learning and strategy pattern recognition involves considering the accurate classification rate (classification accuracies) of the systems or their complement error rates. But, that presumes that each class’s natural distribution (previous probabilities) are balanced and known. In imbalanced settings, in which the positive classes’ prior probability is much lesser in comparison to the negative classes’ (the percentage of these being expressed as the skew), their accuracies are considered insufficient as performance measures because it gets biassed toward the majority classes. This means that, as the skews increase, accuracies tend toward majority classes performance, essentially overlooking the recognition capabilities with regard to the minority classes. That is a rather general (if not the default) occurrence in facial expression recognition settings, in which the prior probabilities of individual target classes (some facial expressions) are considerably lesser in comparison to the negative classes (every other facial expression) (Malathi *et al*., 2019). Therefore, in evaluation of automatic

FER performance, some performance metrics like accuracy (which shows the probability of suitably identifying positive training samples and is independent of prior classes), recall (which signifies the rate of the detected positives which are really precise and, as it merges results from both negative and positive examples, its class prior reliant), F1-measures (which is evaluated as 2 x precision x recall /(precision plus recall)), and ROC (whose calculation involves P(x|positives)/P(x|negatives), whereby P(x|C) indicates the conditional probabilities that information entries have the C class label, and whereby ROC curves plot the result of the classifications to the very negative from the most positive classifications) are much suitable. Moreover, since confusion matrixes present all of the data about the performance of classifiers, it is supposed to be utilised at any time possible to indicate the calculated facial expressions recogniser performance.

Application

The prospective advantages from works in automating the analyses of facial features are diverse and many and encompass a range of fields as diverse as security, education, communication, medicine, and cognitive sciences (Tcherkassof & Dupré, 2020).

In computing technologies and computer science in general, facial expressions offer ways of communicating basic information regarding the demands and needs of the machine. Where the users look (that is, gaze tracking) can be efficiently utilised for freeing computer users from the traditional mouse and keyboard. Moreover, some of the facial gestures (such as winking) can be linked to specific commands (such as mice clicks) providing alternatives to classic mouse and keyboard commands. The human capabilities of ‘‘hearing’’ in a noisy environment via lip reading is the base for bimodal (audio visual) speech processing, which can make the realisation of robust speech-driven user interface systems a reality. To create realistic talking heads (avatars) which represent real individuals, recognising the facial gestures of the individuals and making the avatars take actions based on the facial expressions and synthesised speech is key. Linking facial expression recognisers with facial expression interpretations in expressions and labels such as ‘‘approves’’, ‘‘inattentive’’, ‘‘disagree’’, and ‘‘did not understand’’ can be used as a tool to monitor human reactions in automated tutoring sessions, web-based lectures,

and videoconferences, among others. The somewhat recently introduced affective computing study area focuses majorly on sensing, detection, and interpretation of human affective conditions (like confused, irritated, and pleased, among others) and developing suitable means to handle such affective information in efforts to improve the existing HCI (human-computer interaction) designs. The implied hypothesis is that in most cases human-computer interactions can be enhanced by using machines capable of adapting to their user tendencies as well as their feelings. Seeing that facial expressions is the human direct and natural ways to communicate emotion, machines' analyses of facial expression form a crucial element of affective Human-Computer Interaction design (Li *et al*., 2021).

Facial expressions monitoring and interpretation can also be helpful in terms of providing key information to intelligence and security agents, police, and lawyers about the identity of an individual (some of the works in psychology suggest that facial expression recognitions are much simpler in familiar individuals since it was determined that individuals exhibit similar, ''typical'' facial behaviour patterns in similar situations), deceptiveness (pertinent research in psychology also suggests that visual facial expression features can be used as indications of deception), and attitude (social signals such as mirroring and accord – mimicking postures and facial expressions, among others of the interactions of the partner – are common, basically unconscious signals of needing to be attuned to and be liked by the corresponding partners). What's more, monitoring automated facial reactions can be helpful in law enforcement, as today only informal analyses are basically utilised (Li *et al*., 2021). There are also some systems capable of recognising friendly faces or, more significantly, recognising aggressive or unfriendly ones and informing the appropriate authority, which represent other key applications of facial recognition technology.

Discussion and comparison

The texts presented in this section highlights the significant related works by different scholars in facial expression recognition, including via deep learning and machine learning, among others over recent years. In general, most automatic facial expression

recognition tasks go through various steps such as processing of data, proposed model architectures, and lastly facial recognition.

The pre-processing is a key stage in the process, as indicated by almost all the works reviewed in this sections. The step comprises of a number of techniques like resizing and cropping images to cut down the amount of training, data augmentation, intensity pixels, and normalisation spatial in a bid to increase the images' diversity and get rid of the problem of over-fitting. Each one of these techniques is well articulated by Li *et al.* (2021).

Majority of the methods and systems evaluated here are characterised by high accuracy. Mollahosseini *et al*. (2016) shows how performance can be enhanced by introducing foundation layers in the networks. Chen *et al.* (2020) prefer extracting Action Units from the faces instead of the directly classifying the emotions, Chah *et al.* (2019) is more focused on the issue of occlusion images, also for going into the networks, Fong *et al*. (2021) recommend the addition of the remaining blocks. Namba *et al.* (2020) presents the benefit of addition of the icon face features into the network input and enhanced training using raw images. Laddha & Kumat (2022) presents 2 new CNN architectures after comprehensive analysis of the impacts of convolutional neural network parameters on the accuracy rate of recognition. Majority of the presented methods' results exhibit more than 90 percent accuracy rate.

In general, the research shows that high precision in facial expression recognition can be achieved by the application of convolutional neural networks or related algorithms with spatial data. For sequential data, most of the analysed works use the combination of convolutional neural network and recurrent neural network, particularly long short-term memory network, which shows that convolutional neural network is the basic deep learning network used for facial expression recognition. For the convolutional neural network parameters, the Adam optimization algorithm and Softmax function are the typically utilised. It can also be noted that, for testing the efficiency of the utilised neural network architectures, scholars train and test their models in various databases. The rate of recognition accuracy differs from one database to another using the same deep learning models.

Chapter 3: Requirement Analysis and Methodology

The works presented in this paper analyses the various Facial Expression Recognition (FER) techniques by testing and comparing them against one another to choose the best Deep Learning Algorithm to inject into Android App. To achieve that, the project utilises different FER (facial expression recognition) approaches and methods. In this, five of these methods were implemented them on a computer and were compared based on the following metrics:

- Model: What kind of model/algorithm is used as backbone?
- Dataset: Which dataset is used for training?
- Loss function: What type of loss function is used and why?
- EPOCH: For how many epoch models were trained?
- Face detection: What type of face detector is used in each method?
- Accuracy: % (percentage) of accuracy on each test set.
- Speed: Processing speed on the different machines.

The models were tested on two machines; one with GPU (graphic processing unit) and another with CPU (central processing unit).

This was done in a bid to find the best available solutions in the current technology of this field. By comparing the five FER approaches and methods on the basis of the aforementioned 7 metrics, the best possible option would be determined and be incorporated in a mobile application for the best user experience. In particular, it would result in an easy to use, stable android app that allows even a novice user or non-technical individuals to run and interact with the facial expression recognition (FER) on users' phone without any complex preparations.

Although machine learning approaches could have been also utilised in the project, deep learning techniques were preferred over them for a number of reasons. For instance:

- Human Intervention – machine learning necessitates more constant human interventions to obtain the desired results. On the other hand, while deep learning can be seen as more complex in setting up, it entails minimal intervention after that.

- Hardware – Although machine learning techniques are basically less complex in comparison to deep learning algorithms and are capable of running on standard computers, deep learning algorithms entail far more powerful resources and hardware as well as GPUs (graphical processing units). The utilisation of graphical processing units is beneficial due to their powerful bandwidth memory and abilities of hiding latencies (delays) in memory transfer because of thread parallelism (the capability of numerous operations to efficiently run simultaneously).
- Time – although setting up and operating ML systems can be done faster, they are more limited in the power of their result as compared to deep learning processes which can take up more time in setting up but are capable of generating instant results. Moreover, the quality of the results generated by deep learning algorithms improve with time with more availability of data.
- Approach – ML methods tend to include structured data and utilise traditional algorithms such as linear regression while DL utilises more innovative processes such as neural networks and has the capability of accommodating vast amounts of unstructured data.
- Applications – while ML is used in simple applications such as email inboxes, and doctor's offices, DL algorithms are utilised in and facilitate more autonomous and complex programs, such as robots capable of performing advanced surgery and driverless vehicles, among others.

In particular, a search of the internet resources was done to find out what possible solutions to this problem are available, which resulted in a lot of information, approaches, and techniques to solve the problem of facial expression recognition. As aforementioned, the top five from these methods were chosen and implemented one by one on a local personal computer to test them on public datasets and custom images to determine how good and efficient they would if used in a real world application.

The following subsections describe all the methodologies that were implemented on the two computers (GPU and CPU), including comprehensive details of the implementation. Various algorithms, namely FER 1, FER 2, FER 3, DAN (Deep Attention Neural Network), and RMN (residual masking network).

Chapter 4: Design of Artefact

FER 1

This approach is really remarkable because it provides real-time speed along with satisfactory accuracy. In this approach, the dataset is selected very wisely. The training and evaluation of the model are done using the FER 2013 dataset. This dataset is publicly accessible and is available in two versions. The compressed version of the dataset, which takes 92 megabytes of storage space and the uncompressed version of this dataset, which takes 295 megabytes of storage space. The dataset was split into two subsets. The training set contains 28 thousand images and the testing set contains 3 thousand images in it. The size of each image is uniform with every image in the dataset containing a width of 48 pixels and a height of 48 pixels.

Figure 5: images from the FER 2013 database

As previously noted, deep learning is dominating field in the area of computer vision. Most of the problems in computer vision are actually solved by utilisation of deep learning effectively and efficiently. As such, to solve some of the problems of facial expression recognition such as more time consumption and reliability, deep learning is used. More specifically, a convolutional neural network is used in this context. The CNN is constructed with Keras, which is an interface that uses the TensorFlow backend.

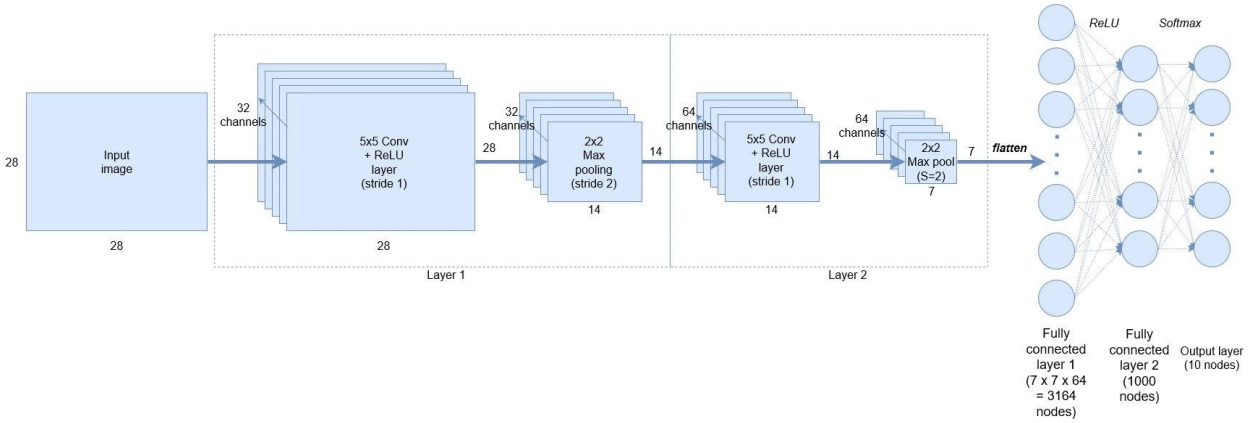


Figure 6: An illustration of convolutional neural network (CNN)

In all the convolution layers, the activation function 'relu' is used while in the dense layer, the activation function 'soft max' is used. To ensure that efficient utilisation of hardware resources for less time, trained set instances which are selected randomly were used. In addition, the loss function “cross-entropy” was used because the problem is multiclass classification.

All the images in the dataset are cropped and only the face portion of each image is presented. In testing the model on custom images, a face detector classifier named “haar cascade” was used. In this, the facial portion of the image was cropped using this detector and then fed to the model to get better results. Caution was taken not to increase the epoch, as that would result in the issue of overfitting; as a result, the epoch was kept small.

FER 2

In this implementation, most of the things are the same that are mentioned in the FER1 but the one important thing that is worth mentioning is that it provides two options for face detection. Based on the use case, either “haar cascade” or “MTCNN” can be used. If the model is used in a situation where speed is the primary objective then “haar cascade” is used but if the main focus is accuracy then “MTCNN” must be used. A tradeoff between speed and accuracy is basically made in this scenario.

This approach is developer-friendly, for instance, it is structured in such a way that the user can get many of the features using very few lines of code, its accuracy is a little bit better than the FER1 although the speed is relatively poor. In all the convolution layers, the activation function 'relu' is used while in the dense layer, the activation function 'soft max' is used. Similarly to ensure efficient utilisation of hardware resources for less time, trained set instances that are selected randomly are used. Also, the loss function “cross-entropy” is used because the problem is multiclass classification.

In this approach, the dataset is selected very wisely. The training and evaluation of the model are done also using the FER2013 dataset.

The OpenCV version must be newer than 3.2 to successfully run this implementation and the Tensorflow must be newer than 1.7.0 for it. In all the convolution layers, the activation function 'relu' is used and in the Dense layer, the activation function 'soft max' is used as well as trained set instances that are selected randomly. The loss function “cross-entropy” is used because the problem is multiclass classification.

FER 3

This includes a PyTorch implementation to solve the issue of the FER. It also uses deep learning to solve the problem. The main model that is used in this implementation is VGG19, which is the most commonly used version of the VGG model. It includes a total

of 19 layers and 16 convolution layers which include 3 fully connected layers in it. All the layers of VGG19 are mentioned in the figure below in column E.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 8: Layers in VGG19

The model was trained on two different datasets (FER2013 and CK+). (The FER2013 dataset is already discussed in the above-mentioned implementation). For a better understanding, the following paragraphs explains the used CK+ dataset.

CK+ dataset

Although the actual name of the dataset is the Extended Cohn-Kanade dataset, it is popularly known by the name the CK+ dataset. There are 593 total videos in the dataset and 123 total different objects. It contains videos of young and old people belonging to different genders and countries. There is a facial expression shift from a neutral expression to a certain expression at a peak level in each video. All the videos are recorded at the rate of 30 FPS. While most of the videos have a resolution of 640x490, some have a resolution of 640x480. The dataset is particularly known for its extensive use in research and development in academics as well as in the industry. All the other metrics in this dataset are the same as are discussed in the FER1. For instance, to ensure that efficient utilisation of hardware resources for less time, trained set instances which are selected randomly were used and the loss function “cross-entropy” was used because the problem is multiclass classification

Figure 9: Extended Cohn-Kanade dataset

As afore-mentioned, the model is trained on two different datasets. The dataset was used because the accuracy of the model trained on CK+ is really remarkable although the model trained on FER2013 is also not very poor. This model yields really exceptional speed when used in GPU. (The speed and performance of each model as well as the comparisons are presented in the following results chapter).

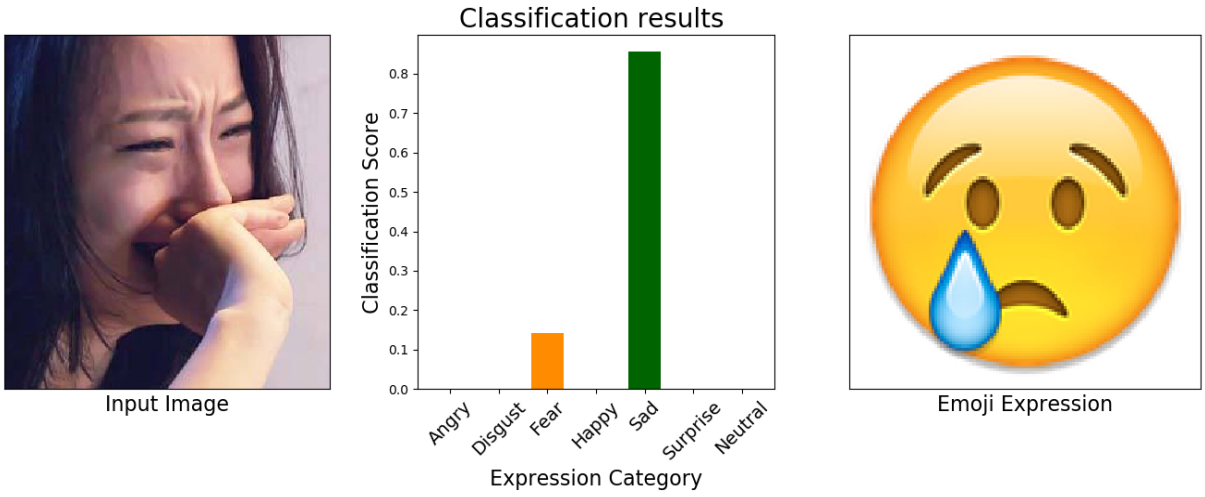


Figure 10: CNN based PyTorch implementation on facial expression recognition (FER2013 and CK+)

DAN

Distract Your Attention: Multi-head Cross Attention Network for Facial Expression Recognition

This implementation was done using PyTorch similarly to the aforementioned processes. The main network that was used to train the model is resnet18, which is a convolutional neural network that consists of 18 layers. Figure 11 below illustrates the architecture of RESNET18:

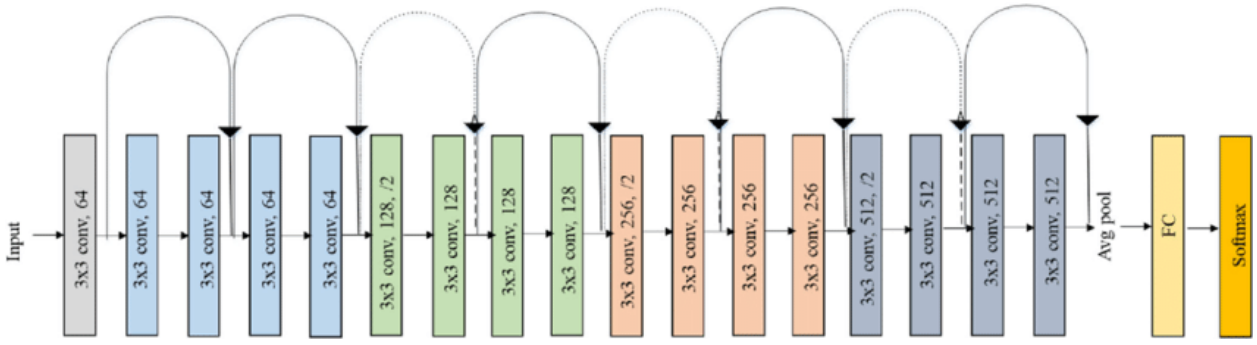


Figure 11: RESNET 18 architecture

A combination of datasets were used to train, evaluate and test the trained model in this case. The first dataset that was used for training the model is known as AffectNet, which is one of the world's largest datasets with 400 thousand labelled images. All the images were labelled manually and the labels present the expression on the face in the particular image. It also provides the intensity of various expressions in different images.

Moreover, Real-world Affective Faces Database (RAF-DB), one of the large scale datasets related to the FER (facial expression recognition) was used. The dataset contains 30 thousand images with an extensive variety. All the images are downloaded from the internet and the dataset is annotated using crowdsourcing. In other words, each image in the dataset is annotated independently by more than 40 annotators.

The other metrics used in this case are similar to the ones mentioned in the FER1. Its speed was found to be higher but the model is not much accurate as the aforementioned FER3. What is more, it is very fast when GPU is used to run it. The main problem, however, is that the face detection used in this approach is not very efficient.

RMN (Residual Masking Network)

Facial Expression Recognition using Residual Masking Network with PyTorch

A PyTorch implementation was also utilised in this case with the main network used in this implementation being res10_300 x 300 SSD. It is a variant of RESNET but with the single-shot detection technique incorporated in it. It is much more accurate than the other implementations other than FER3. Its accuracy is really good although it compromises on the speed while using this algorithm.

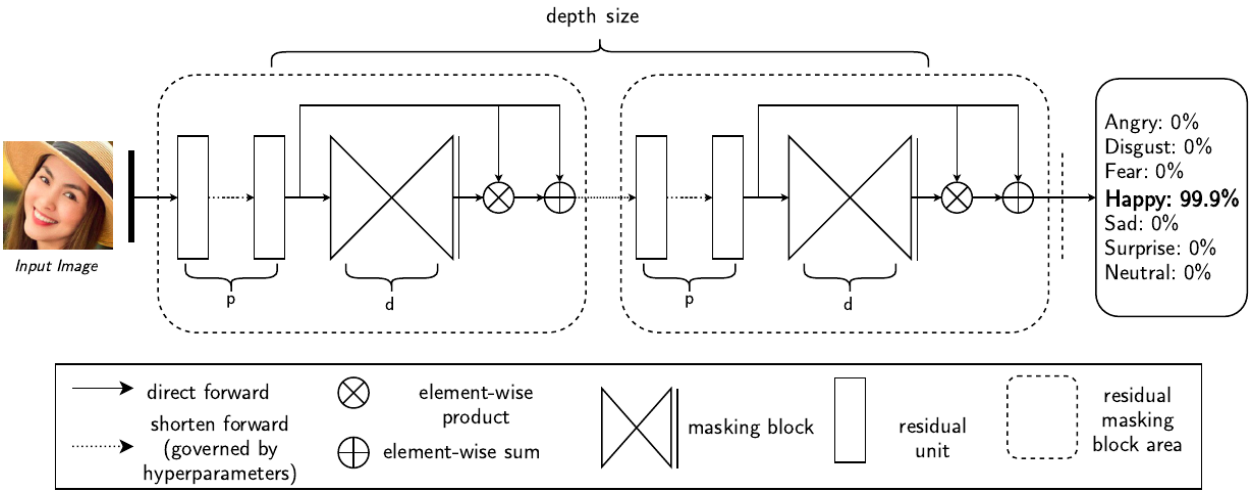


Figure 12: Basic architecture of Residual Masking Network

In this implementation, again, a combination of two datasets is used to train the model. The first one is FER2013, which was discussed earlier in detail, and the other one is known as the VEMO dataset.

In particular, the face detector used in this technique is the SSD face detector, which is a really fast and fairly accurate deep learning-based object detection method. In our case, it is customised to detect the faces from the given images and follows the exact concept of YOLO. One of the observed things during the implementation is that this face detector was not able to detect faces from animated characters and images or videos where the faces were kind of blurred. As a result, it would skip the face in such instances. This was one of encountered issues during the testing of this approach. To deal with, a custom solution was applied by adding a few lines of code to create the logic that if there is no face detected in the image then that particular image would be deleted and the process would be resumed.

The loss function “cross-entropy” was used because the problem is multiclass classification and the model was trained for 50 epochs in this approach. There was no overfitting problem in the model when it was tested on custom images.

Results

During the test, public datasets or images online were particularly used for testing each of the algorithm and environment.

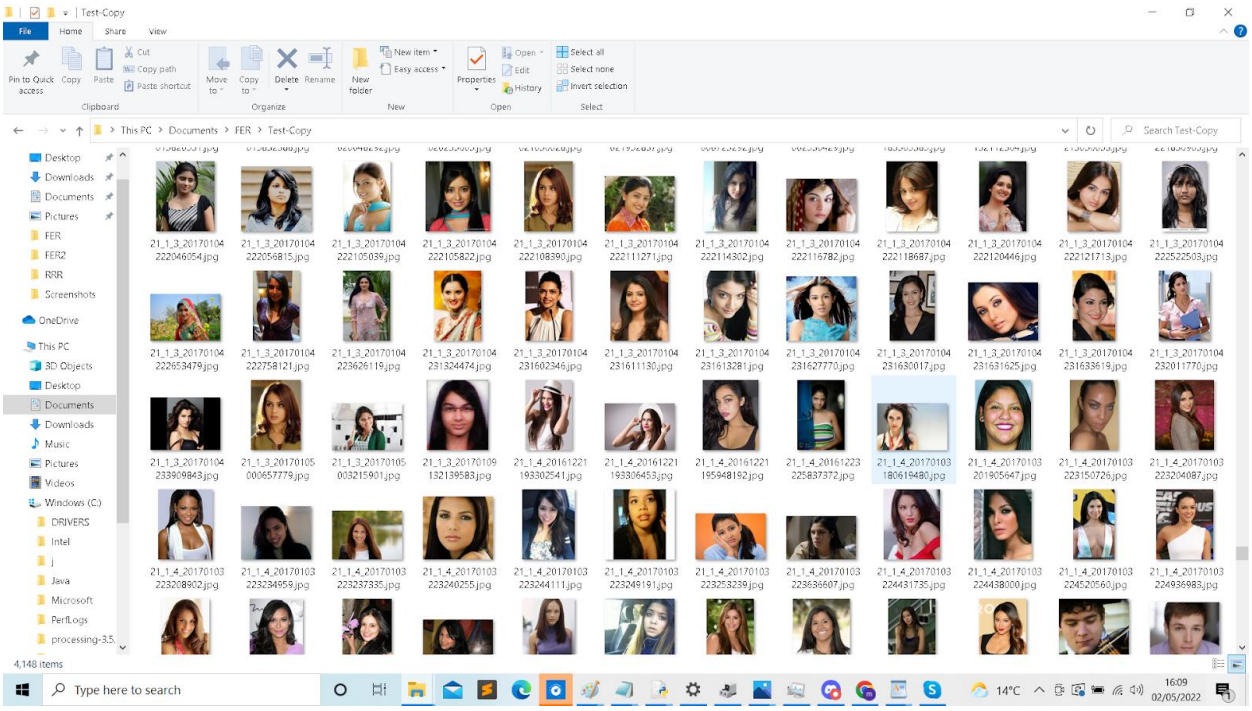


Figure 13: Example of the public dataset for images used in the test

Figure 13 above includes an example of the used public dataset for images. The used images were specifically open-source and free to use and hence, were used for testing on the models. Caution was taken to ensure that the images were unique as the model was trained using private dataset. As a result, using the public dataset would make it lose some of the speed and accuracy as it gets the new information that it was fed.

The following figures present the testing speed and other results obtained for using the models on the pc version:

Figure 14: FER 2 in the PC version

In FER2 (facial, expression recognition), the test running on the laptop was found to be light and did not use an intensive amount of GPU (it only used about 1-5% of the GPU) while it used 100% of the CPU.

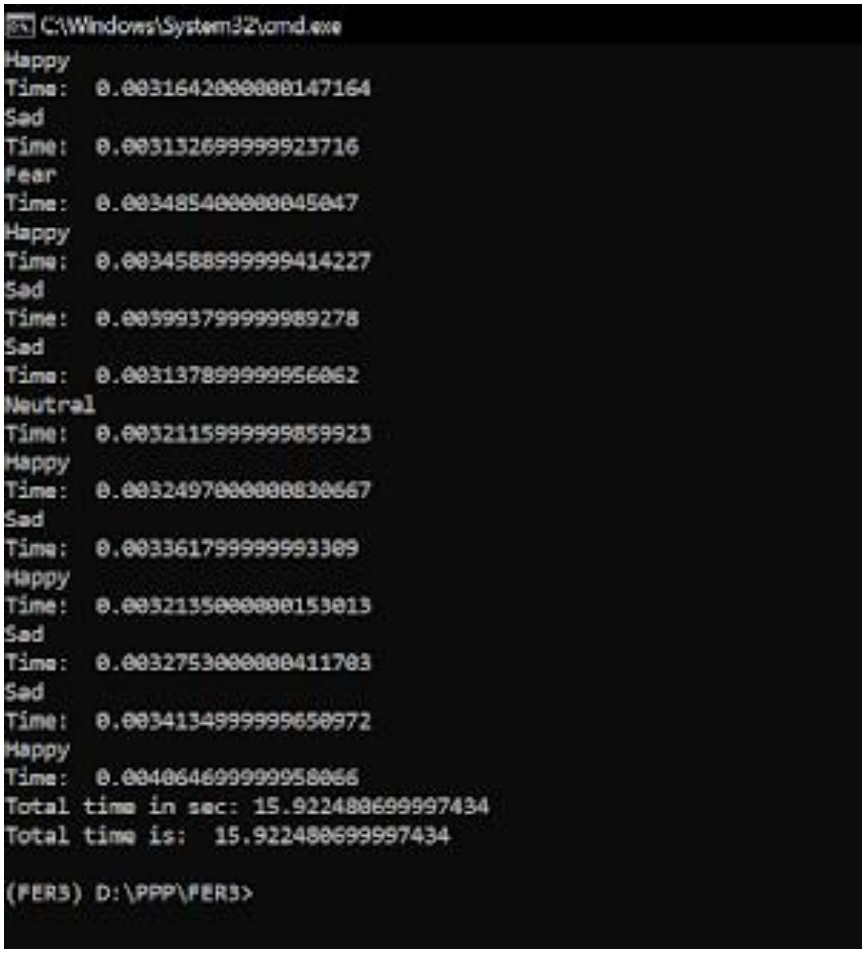


Figure 15: Results of FER 3 on a PC

In the figure above, 15 seconds were taken to use 100% of CPU and 40+ GPU with no background software such as google chrome and other background application..

This was determined to be the fastest and most accurate model while being used on a PC with a good GPU. During testing, the GPU percentage was recorded above 60% which is why its speed got to a total of 15 seconds.

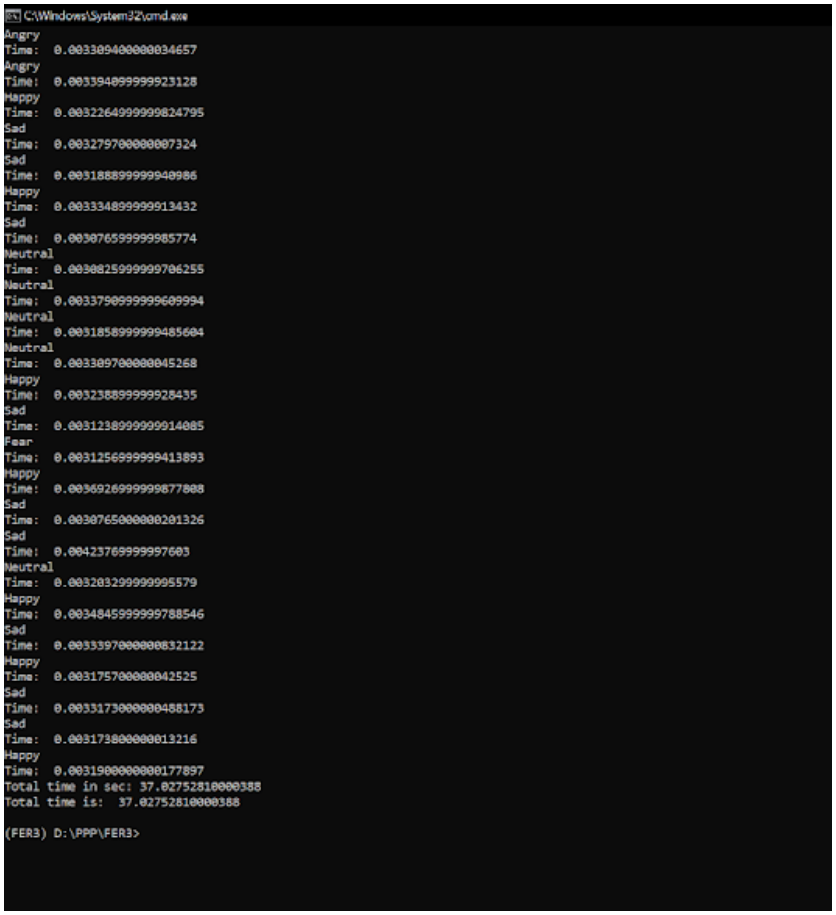


Figure 15: Results of FER 3 on a PC

The test was run again using Google chrome and other web browsers and with the time increased to 37 seconds. In this time, a software was running in the background to check whether the speed would be affected if other software takes about 1% or more of CPU and memory, which proved the study theory on testing this.

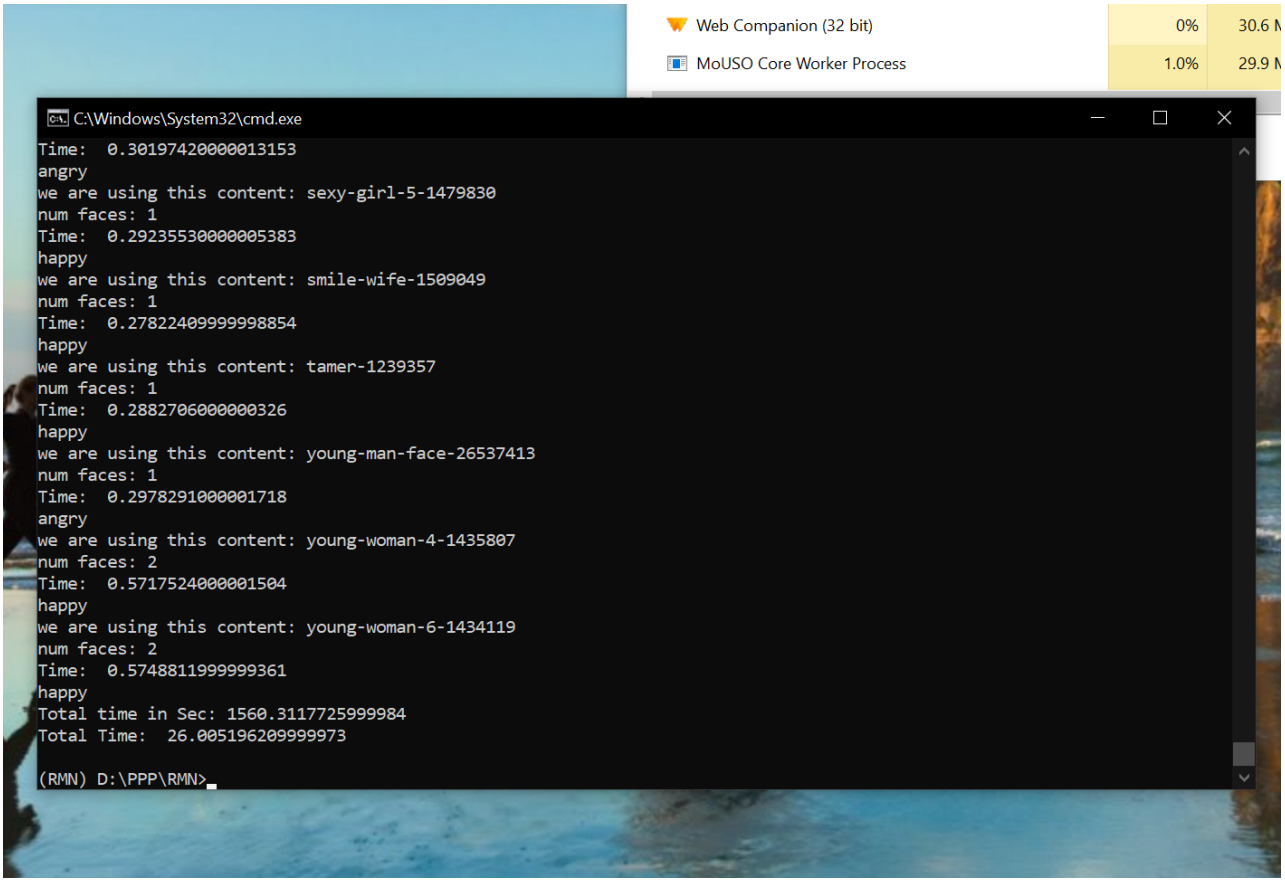


Figure 16: The results of RMN

One of the main issues with RMN is that the accuracy was found to be very poor and sometimes could not even detect the face due to light and angle. For instance, in the image below, the person in the image shows a happy facial expression while the RMN testing indicated that the person was displaying an angry facial expression.

The mechanism of face detection in this implementation is not very fine. It ended up failing to recognise any face, that is, the model for face detection could not detect the human face due to the angle or lighting and also the facial expression. Besides, the model could also not recognise any images with a ".webp" extension when used as a substitution of jpeg and/or png.

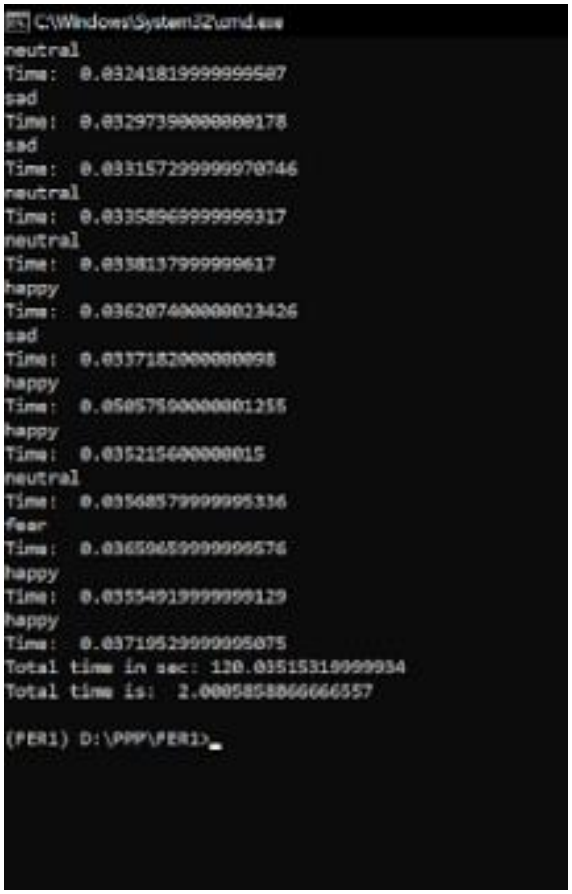
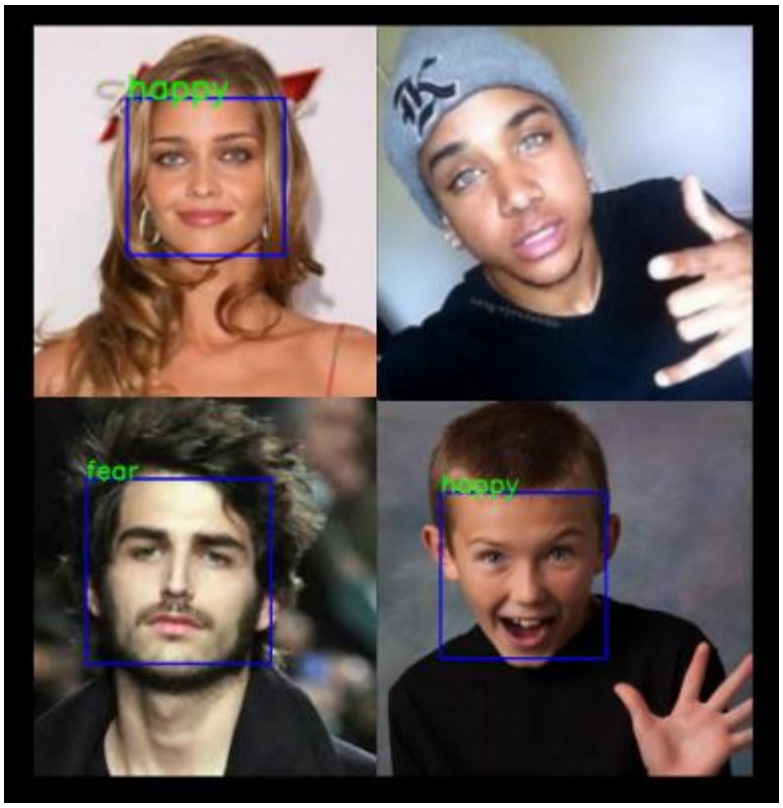


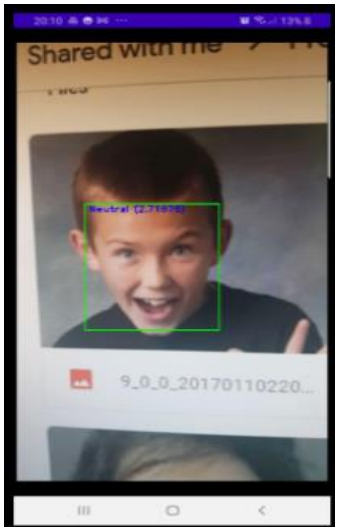
Figure 17: FER 1 results from the test

After the test on this model, the FER 1 was found to include very few minor background errors and the overall speed is 2 minutes for the PC testing stand point. The model was also ran multiple times to check whether if there would be any speed differences but there was no major speed difference that affect the performance in accuracy and speed.

For instance, as seen in the following tested Images, the model could detect some of the faces even with the lighting variations and different angles positions.



In a different angle, however, as shown in the image below, the facial expression displayed is detected wrongly as neutral. This is because FER 1 accuracy is not as good compared to others like FER 3.



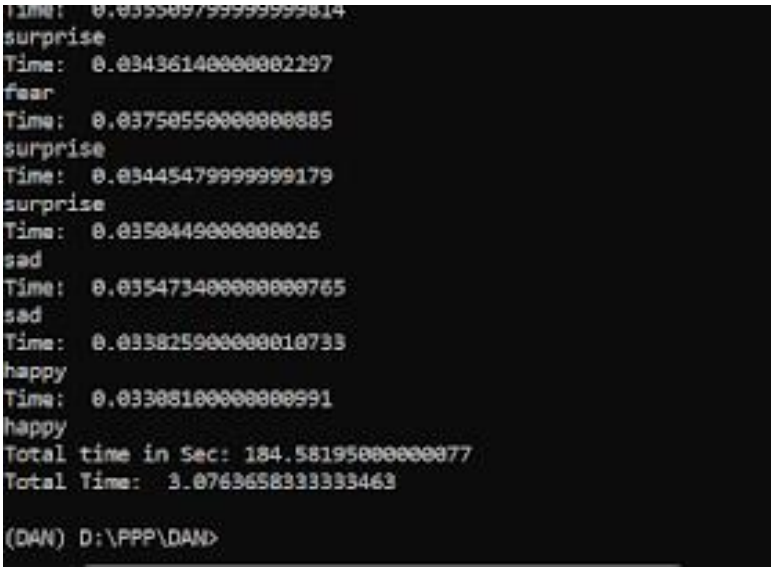


Figure 18: DAN model speed test on PC

Chapter 5: Testing and Evaluation of the Artefact

Laptop Version

During testing on the Laptop version, it was conducted on the 10th gen i5 Intel for each of the facial expression recognition models. The speed difference is very noticeable and in comparison to my PC which has a GPU 2070RTX NVidia and CPU 9th gen i7, it shows how much of a difference there is when using a GPU. It, specifically, helps to solve the valuation at a very fast pace. During the model testing on the laptop, the processor seemed a bit slower as it was heating up every time. As a result, the test run for the very compute intensive models such as RMN. RMN took the longest time as it was trying to increase the accuracy more than speed for its default state.

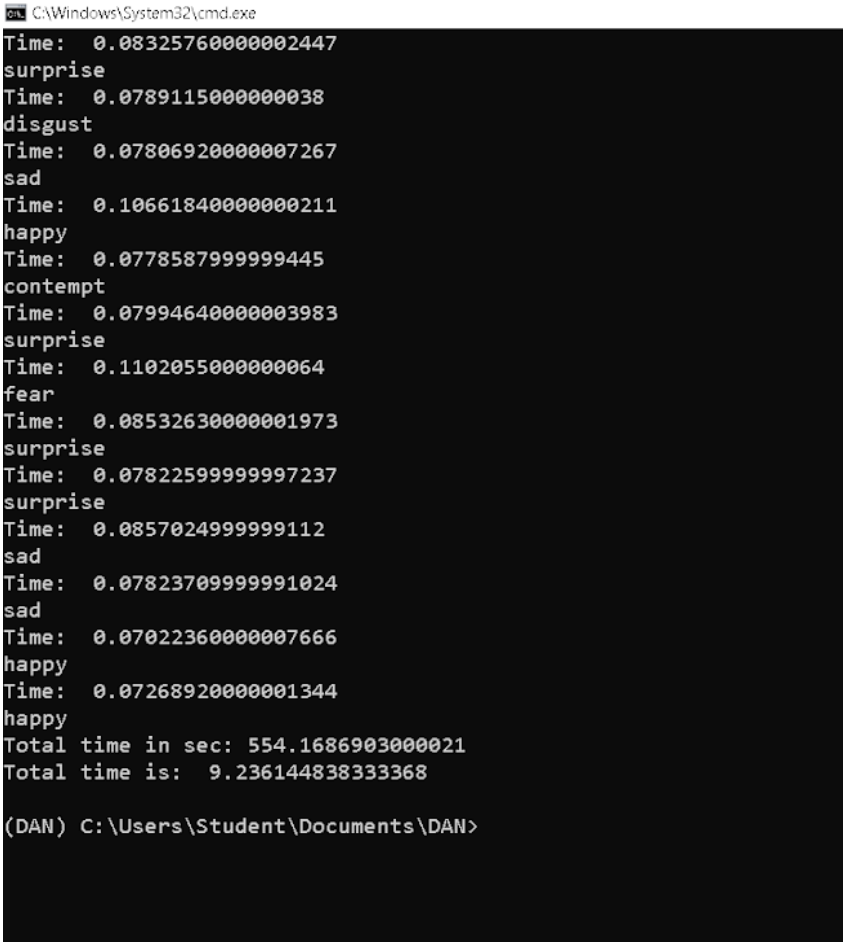


Figure 19: DAN on laptop

The image detection by DAN was the slowest in comparison to the other models and It took 9 minutes and 23 seconds.

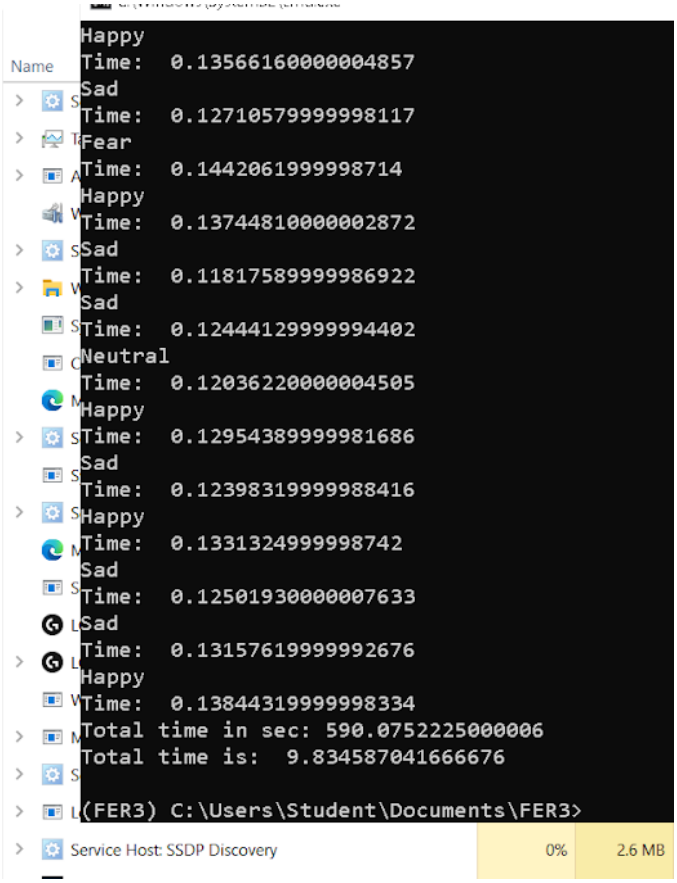


Figure 20: FER 3 on a laptop

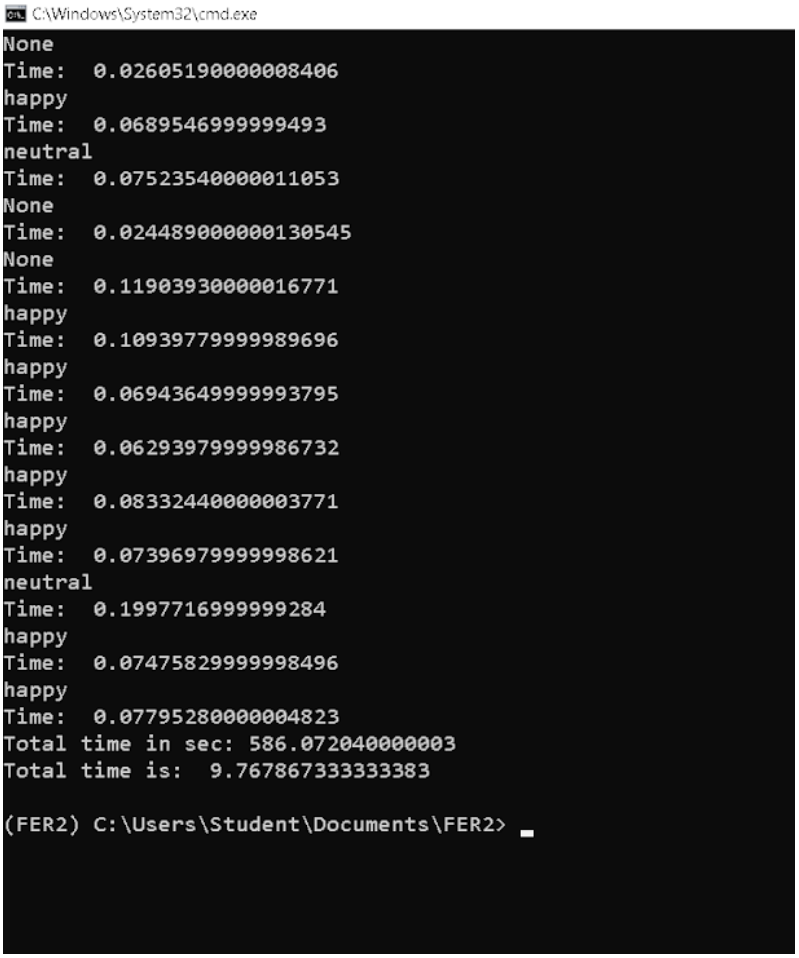


Figure 21: FER2 on laptop

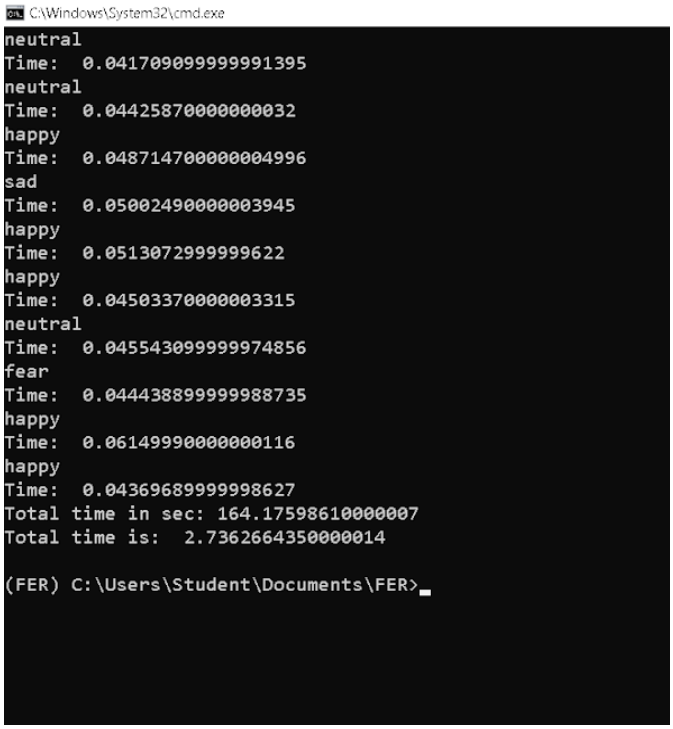


Figure 22: FER 1 laptop

The time taken by FER 1 was between 2 minutes 44 seconds and 2 minutes 58 seconds.

This testing speed had some slight changes due to the software running on the background.

Overall comparison

RMN on laptop had no software running in the background that would be taking up some of the CPU processing power. In the first attempt, it took 50min 20secound while in the second attempt, it was between 48min & 68 sec and 49 minutes & 8 seconds. Following that, a third test attempted to see if the speed would reduce or increase. It was found that the speed run time results of the first two attempts were similar. When I run this code constantly for a long period of time, laptop is heated so much. With the increase in heat, processing speed is decreased.

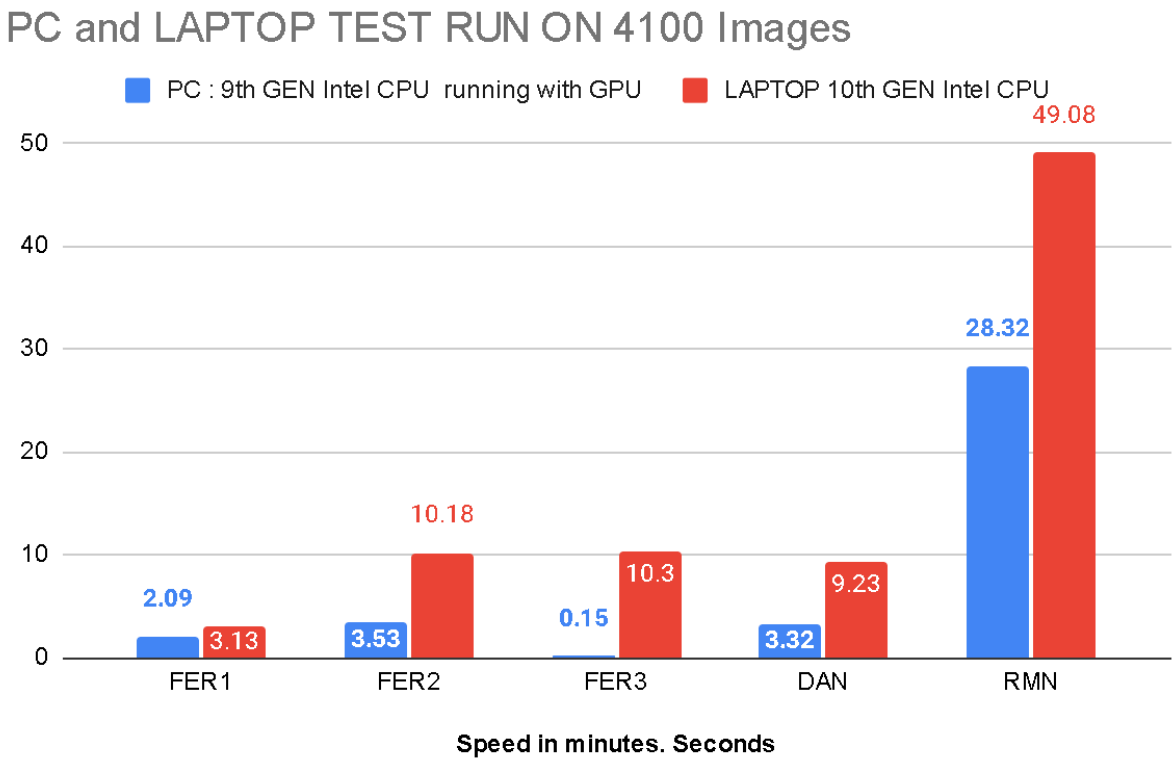


Figure 23: comparison of data for the five models for both laptop and Pc versions

The data presented in the figure above shows the comparison of the five models in laptop (10^{th} GEN Intel CPU) and PC (9^{th} GEN Intel CPU running with GPU of Nvidia 2070 RTX) version. Specifically, the presented data is the speed in minutes and seconds of each model. The result shows that RMN takes the longest with 49.8 minutes in the laptop and 28.32 minutes in the PC, followed by FER 3 with 14.3 minutes in the laptop and 0.15 in the PC, followed by FER 2 with 10.18 minutes in the laptop and 3.53 minutes in the PC, followed by DAN with 6.48 minutes in the laptop and 3.32 minutes in the PC, and lastly is the FER 1 with 2.58 minutes in the laptop and 2.09 minutes in the PC, making it the fastest and most reliable in terms of speed.

In summary, the most appropriate algorithm was found to be the FER1. This is because of a number of reasons, including:

- It is more memory efficient.
- It is power efficient.
- It is not compute-intensive.
- It is a small size model.
- It has fast processing.
- It includes less dependencies.
- It is easy to modify the code in the model.
- It is a light weight model for mobile applications.
- It is fast enough to run in real time.
- It includes satisfactory accuracy with fast speed.

Links of each FER models :

FER 1 - https://sefiks.com/2018/01/01/facial-expression-recognition-with-keras/

FER2 - https://github.com/justinshenk/fer

FER3 Link: WuJie1010/Facial-Expression-Recognition.Pytorch: A CNN based pytorch implementation on facial expression recognition (FER2013 and CK+), achieving 73.112% (state-of-the-art) in FER2013 and 94.64% in CK+ dataset (github.com)

Dan- yaoing/DAN: Official implementation of DAN (github.com)

RMN - phamquiluan/ResidualMaskingNetwork: Facial Expression Recognition using Residual Masking Network (github.com)

Chapter 6: Evaluation and Conclusion of the Project

The works presented in this paper sets to test various facial expression models based on various metrics, the type of algorithm used, the utilised dataset, the employed loss function, the number of trained epoch models, the type of face detector utilised in each method, the percentage accuracy on each test set, and the processing speed on the different computer systems. The specific utilised models are FER 1, FER 2, FER 3, DAN, and RMN. From the experiment, it is evident that there is always a tradeoff between speed and accuracy in facial recognition algorithms. As such, if the best accuracy is what is needed, then they has to be a compromise on speed and so on.

Conclusion

From the experiment and specifically, the conducted tests, FER1 was found to have the best speed in general for both PC and laptop. FER 3, on the other hand, was found to include the best accuracy on using the VGG19 model for PC and laptop. I used FER1 for my mobile application based on the matrices that I mentioned in this report.

Future Work

As a future work, the information and knowledge obtained in this research and analysis can be used to develop a model capable of getting the maximum accuracy without compromising the speed as much as possible. This would be helpful for a number of ways, for instance, the model can be useful in instantly detecting facial expressions and applied in computer systems for recognising and classifying the emotions accordingly. This is because of the high accuracy rate of the model.

References

Black, M., & Yacoob, Y. (2015). Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. *Inproceedings Of The International Conference On Computer Vision*, *25*(3), 297-310. https://doi.org/10.1016/j.imavis.2005.10.004

Brown, F. (2020). Facial Expression Detection using Artificial Intelligence. *International Journal Of Recent Technology And Engineering*, *8*(5), 1720-1723. https://doi.org/10.35940/ijrte.e6284.018520

Kouda, T. (2018). Detection of Blink and Facial Expression Changes using DCT Signs. *Journal Of The Institute Of Industrial Applications Engineers*, *2*(2), 70-73. https://doi.org/10.12792/jiiae.2.70

Liu, Y., Li, Y., Ma, X., & Song, R. (2017). Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas. *Sensors*, *17*(4), 712. https://doi.org/10.3390/s17040712

Namba, S., Sato, W., Osumi, M., & Shimokawa, K. (2021). Assessing Automated Facial Action Unit Detection Systems for Analyzing Cross-Domain Facial Expression Databases. *Sensors*, *21*(12), 4222. https://doi.org/10.3390/s21124222

Shaham, G., Mortillaro, M., & Aviezer, H. (2020). Automatic facial reactions to facial, body, and vocal expressions: A stimulus-response compatibility study. *Psychophysiology*, *57*(12). https://doi.org/10.1111/psyp.13684

Andre, E. (2021). Editorial: Transactions on Affective Computing – Affective Computing in the Times of Pandemics. *IEEE Transactions On Affective Computing*, *12*(1), 1-1. https://doi.org/10.1109/taffc.2021.3059491

Filippini, C., Perpetuini, D., Cardone, D., & Merla, A. (2021). Improving Human–Robot Interaction by Enhancing NAO Robot Awareness of Human Facial Expression. *Sensors*, *21*(19), 6438. https://doi.org/10.3390/s21196438

Filko, D., & Martinović, G. (2013). Emotion Recognition System by a Neural Network Based Facial Expression Analysis. *Automatika*, *54*(2), 263-272. https://doi.org/10.7305/automatika.54-2.73

Maroti, D., Ljótsson, B., Lumley, M., Schubiner, H., Hallberg, H., Olsson, P., & Johansson, R. (2021). Emotional Processing and Its Association to Somatic Symptom Change in Emotional Awareness and Expression Therapy for Somatic Symptom Disorder: A Preliminary Mediation Investigation. *Frontiers In Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.712518

Mellers, B., Schwartz, A., Ho, K., & Ritov, I. (2017). Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science*, *8*(6), 423-429. https://doi.org/10.1111/j.1467-9280.1997.tb00455.x

Namba, S., Sato, W., Osumi, M., & Shimokawa, K. (2021). Assessing Automated Facial Action Unit Detection Systems for Analyzing Cross-Domain Facial Expression Databases. *Sensors*, *21*(12), 4222. https://doi.org/10.3390/s21124222

Picard, R., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, *23*(10), 1175-1191. https://doi.org/10.1109/34.954607

Suwa, M., Sugie, N., & Fujimora, K. (2008). A preliminary note on patternrecognition of human emotional expression. *International Joint Conference On Pattern Recognition*, 408-410. Retrieved 25 April 2022, from.

Barbosa, J., Seo, W., & Kang, J. (2019). paraFaceTest: an ensemble of regression tree-based facial features extraction for efficient facial paralysis classification. *BMC Medical Imaging*, *19*(1). https://doi.org/10.1186/s12880-019-0330-8

Chah, N. (2019). Down the deep rabbit hole: Untangling deep learning from machine learning and artificial intelligence. *First Monday*. https://doi.org/10.5210/fm.v24i2.8237

Ekman, P. (2017). Emotional and Conversational Nonverbal Signals. *Language, Knowledge, And Representation*, 39-50. https://doi.org/10.1007/978-1-4020-2783-3_3

Fong, A., & Hong, G. (2021). Boosted Supervised Intensional Learning Supported by Unsupervised Learning. *International Journal Of Machine Learning And Computing*, *11*(2), 98-102. https://doi.org/10.18178/ijmlc.2021.11.2.1020

Fourie, C. (2003). Deep learning? What deep learning?. *South African Journal Of Higher Education*, *17*(1). https://doi.org/10.4314/sajhe.v17i1.25201

Hameed, S. (2014). Facial Features Extraction by Relative Geometrical Position. *International Journal Of Computer Applications*, *98*(15), 37-40. https://doi.org/10.5120/17263-7619

Laddha, S., & Kumar, V. (2022). DGCNN: deep convolutional generative adversarial network based convolutional neural network for diagnosis of COVID-19. *Multimedia Tools And Applications*. https://doi.org/10.1007/s11042-022-12640-6

Malhotra, Y. (2018). AI, Machine Learning & Deep Learning Risk Management & Controls: Beyond Deep Learning and Generative Adversarial Networks: Model Risk Management in AI, Machine Learning & Deep Learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3193693

Mark, E. (2017). Neural Networks. *Neural Networks*, *85*, v-ix. https://doi.org/10.1016/s0893-6080(16)30177-0

Neckel, S., & Hasenfratz, M. (2021). Climate emotions and emotional climates: The emotional map of ecological crises and the blind spots on our sociological landscapes. *Social Science Information*, 053901842199626. https://doi.org/10.1177/0539018421996264

Sáez Trigueros, D., Meng, L., & Hartnett, M. (2021). Generating photo-realistic training data to improve face recognition accuracy. *Neural Networks*, *134*, 86-94. https://doi.org/10.1016/j.neunet.2020.11.008

Sokolov, V. (2017). Discussion of ‘Deep learning for finance: deep portfolios’. *Applied Stochastic Models In Business And Industry*, *33*(1), 16-18. https://doi.org/10.1002/asmb.2228

Solopchuk, O., & Zénon, A. (2021). Active sensing with artificial neural networks. *Neural Networks*, *143*, 751-758. https://doi.org/10.1016/j.neunet.2021.08.007

Sutskever, I., & Hinton, G. (2020). Temporal-Kernel Recurrent Neural Networks. *Neural Networks*, *23*(2), 239-243. https://doi.org/10.1016/j.neunet.2009.10.009

Zhu, G., & Zhao, T. (2021). Deep-gKnock: Nonlinear group-feature selection with deep neural networks. *Neural Networks*, *135*, 139-147. https://doi.org/10.1016/j.neunet.2020.12.004

AL-Oudat, M., Azzeh, M., Qattous, H., Altamimi, A., & Alomari, S. (2022). Image Segmentation based Deep Learning for Biliary Tree Diagnosis. *Webology*, *19*(1), 1834-1849. https://doi.org/10.14704/web/v19i1/web19123

Chen, A., Xing, H., & Wang, F. (2020). A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing. *IEEE Access*, *8*, 49741-49751. https://doi.org/10.1109/access.2020.2980060

Dyrka, W., Pyzik, M., Coste, F., & Talibart, H. (2019). Estimating probabilistic context-free grammars for proteins using contact map constraints. *Peerj*, *7*, e6559. https://doi.org/10.7717/peerj.6559

Fadaei, Y., & Moghadam, M. (2017). Approximate solutions of partial differential equations by some Meshfree Greedy Algorithms. *Numerical Methods For Partial Differential Equations*, *33*(6), 1884-1899. https://doi.org/10.1002/num.22164

Kaur, R. (2022). From machine learning to deep learning: experimental comparison of machine learning and deep learning for skin cancer image segmentation. *Rangahau Aranga: AUT Graduate Review*, *1*(1). https://doi.org/10.24135/rangahau-aranga.v1i1.32

Koroscik, J. (2022). The Effects of Prior Knowledge, Presentation Time, and Task Demands on Visual Art Processing. *Studies In Art Education*, *23*(3), 13. https://doi.org/10.2307/1320012

Mochihashi, D. (2020). Robotics, Grounding and Natural Language Processing. *Journal Of Natural Language Processing*, *27*(4), 963-968. https://doi.org/10.5715/jnlp.27.963

Mollahosseini, A., Chan, D., & Mahoor, M. (2016). Going deeper in facial expression recognition using deep neural networks. *2016 IEEE Winter Conference On Applications Of Computer Vision (WACV)*, 1-10. Retrieved 2 May 2022, from.

Najafian, M., & Russell, M. (2020). Automatic accent identification as an analytical tool for accent robust automatic speech recognition. *Speech Communication*, *122*, 44-55. https://doi.org/10.1016/j.specom.2020.05.003

Pevzner, P. (2019). Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians. *Bioinformatics*, *20*(14), 2159-2161. https://doi.org/10.1093/bioinformatics/bth217

Rupa, G. (2021). Preclinical Pharmacology and Toxicology: an Important Aspect in Drug Discovery. *Advances In Clinical Toxicology*, *1*(1). https://doi.org/10.23880/act-16000101

Schedl, M. (2019). Deep Learning in Music Recommendation Systems. *Frontiers In Applied Mathematics And Statistics*, *5*. https://doi.org/10.3389/fams.2019.00044

Sharma, R. (2020). Study of Supervised Learning and Unsupervised Learning. *International Journal For Research In Applied Science And Engineering Technology*, *8*(6), 588-593. https://doi.org/10.22214/ijraset.2020.6095

Su, L., & Qi, Y. (2018). Software Group Rejuvenation Based on Matrix Completion and Cerebellar Model Articulation Controller. *Neuroquantology*, *16*(5). https://doi.org/10.14704/nq.2018.16.5.1354

Wen, S. (2019). Translation analysis of English address image recognition based on image recognition. *EURASIP Journal On Image And Video Processing*, *2019*(1). https://doi.org/10.1186/s13640-019-0408-9

Appendices

Tutorial : Click on this link on OneDrive to access the project.

click here : Final year project

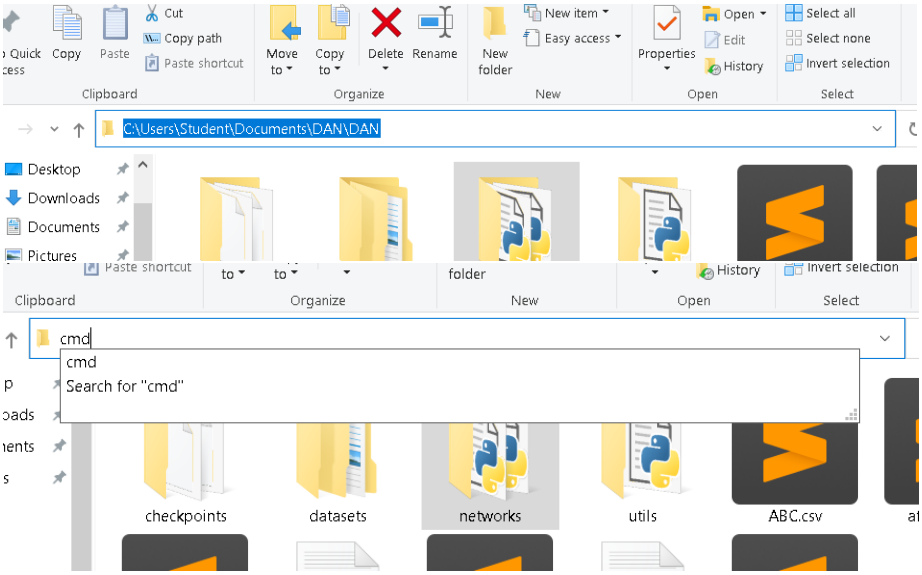
Instructions on how to do the test speed and use the application on your android phone. You must read the help File as this explains what is needed further.

Step 1: Click on the link for OneDrive.

Step 2 : Read the Help file.

Step 3 : After reading the help file and install the model folder you picked.

Then place it in a suitable place in either the c or d drive.



Go to inside the FER model and type cmd.

Step 4 : Then the command prompt will pop up and you will have to type this in :

- pip install -r requirements.txt

```
Microsoft Windows [Version 10.0.19044.1645]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Student\Documents\FER3\FER3>pip install -r requirements.txt
Defaulting to user installation because normal site-packages is not writeable
Collecting cycler==0.11.0
  Using cached cycler-0.11.0-py3-none-any.whl (6.4 kB)
Collecting fonttools==4.32.0
  Using cached fonttools-4.32.0-py3-none-any.whl (900 kB)
Collecting imageio==2.17.0
```

(this has been done above in order to install all the dependencies)

All the testing and development is done using python 3.7.0. If you have another version of python you must delete that. If you don't know how to find and delete that version of python on your computer, go to Google and find a guide on how to delete this.

Step 5 : Download the FER (facial expression recognition) models of your chosen type from the oneDrive which is mentioned above with the link..

If you are using a Nvidia GPU - read this below :

- If you are using a Nvidia GPU and you have not activated it and dont have cuDNN and cuDA installed then follow the instructions on Tensorflow after running the model.
- There were two or three expected to be downloaded and you have to place it in the right directory. Here provide information on fixing the GPU issue
- https://stackoverflow.com/questions/51306862/how-do-i-use-tensorflow-gpu
- https://www.codingforentrepreneurs.com/blog/install-tensorflow-gpu-windows-cuda-cudnn/

If you don't have an Nvidia GPU and have a different GPU, then follow the guide on what Tensorflow tells you by typing in your version type of GPU.

- Then go to oneDrive and
- select a model to download.
- Then double click the file directory and input CMD
- Then the command prompt will show up.
- input in as follow :
    - cd Scripts
    - activate
    - cd ..
    - python Appfinal.py

Then the images should now run.

The “python” at the start needs to be there for it to read any file in the directory. AppFinal.py or anything testfinal.py.

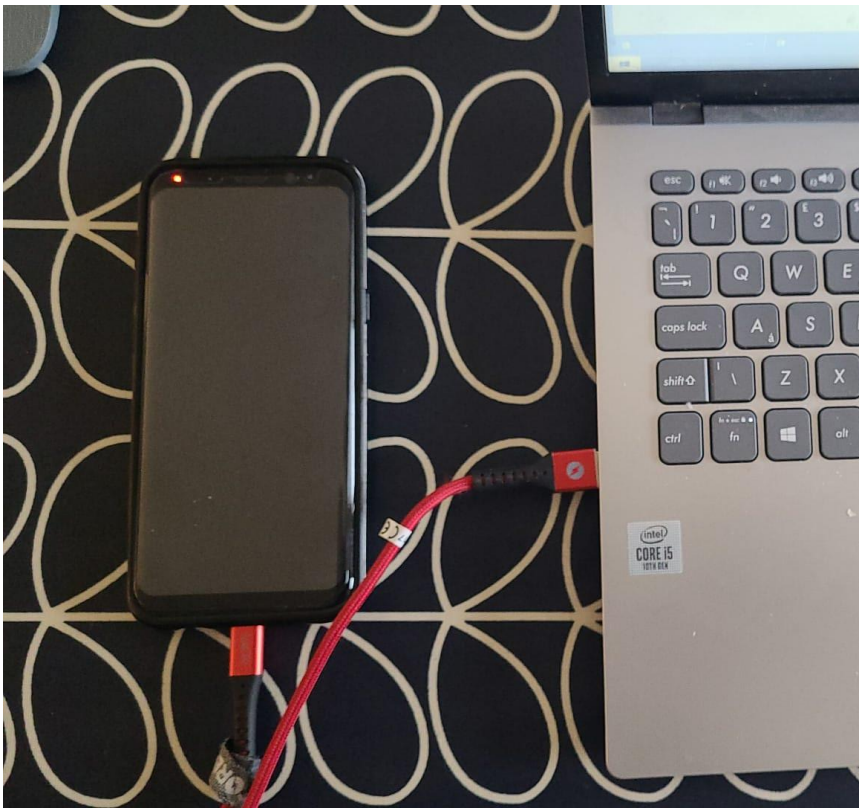
Allow it to run and you get the speed at the end.

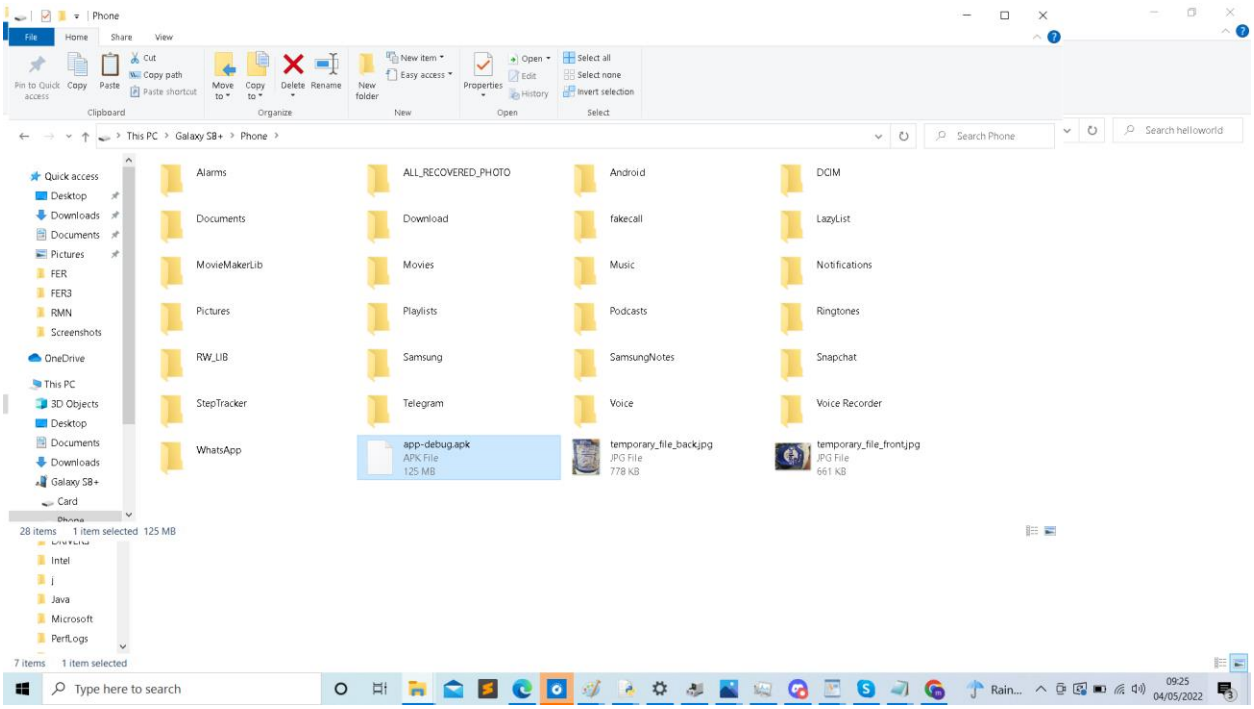
To create the apk file you need to use the Cradle folder that is provided in the mobileapp folder.

Here are the commands you will need to put:

- Create a virtualenv
- Then activate it, I provided steps to do it in the help file
- python -m pip install briefcase
- briefcase new
- “keep all the options default”
- cd helloworld
- briefcase create android
- You will get 9 files here “helloworld\android\cradle\Hello World”
- Delete them all and use the files I provided:
- “use the cradle folder instead of the default folder”
- briefcase build android
- You will the apk file here:
- Built android\gradle\Hello World\app\build\outputs\apk\debug\app-debug.apk
- The file size will be 125mb the ‘fer.apk’.
- Copy the apk file to your mobile and follow the install instruction mentioned below

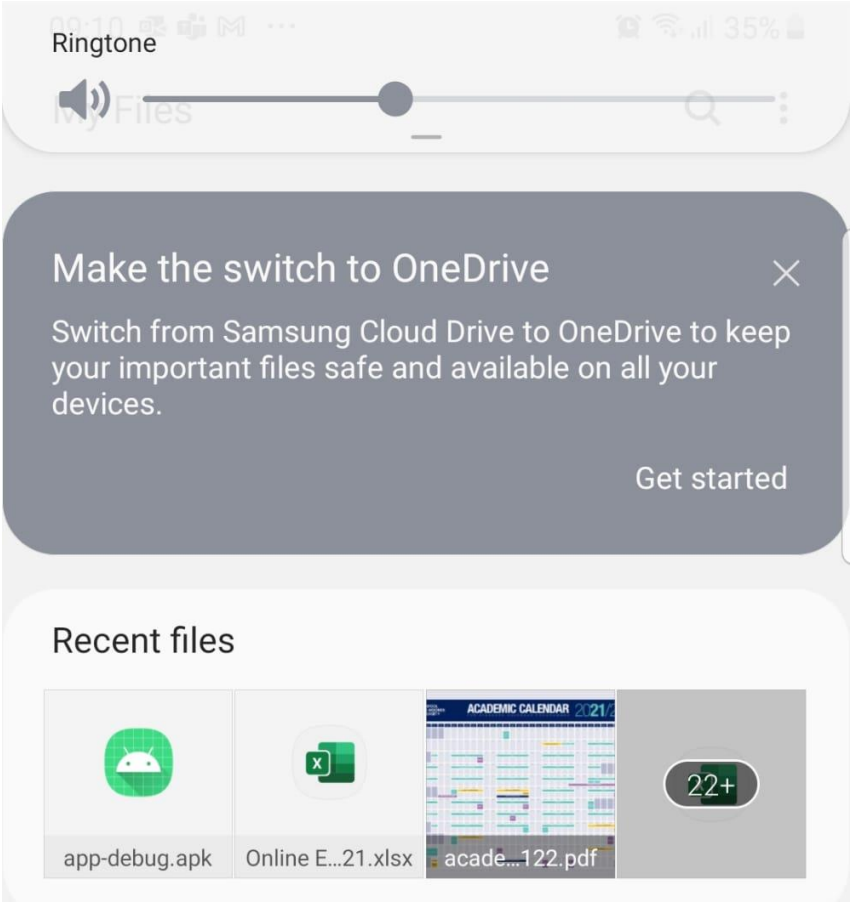
Tutorial for how to use mobile application.





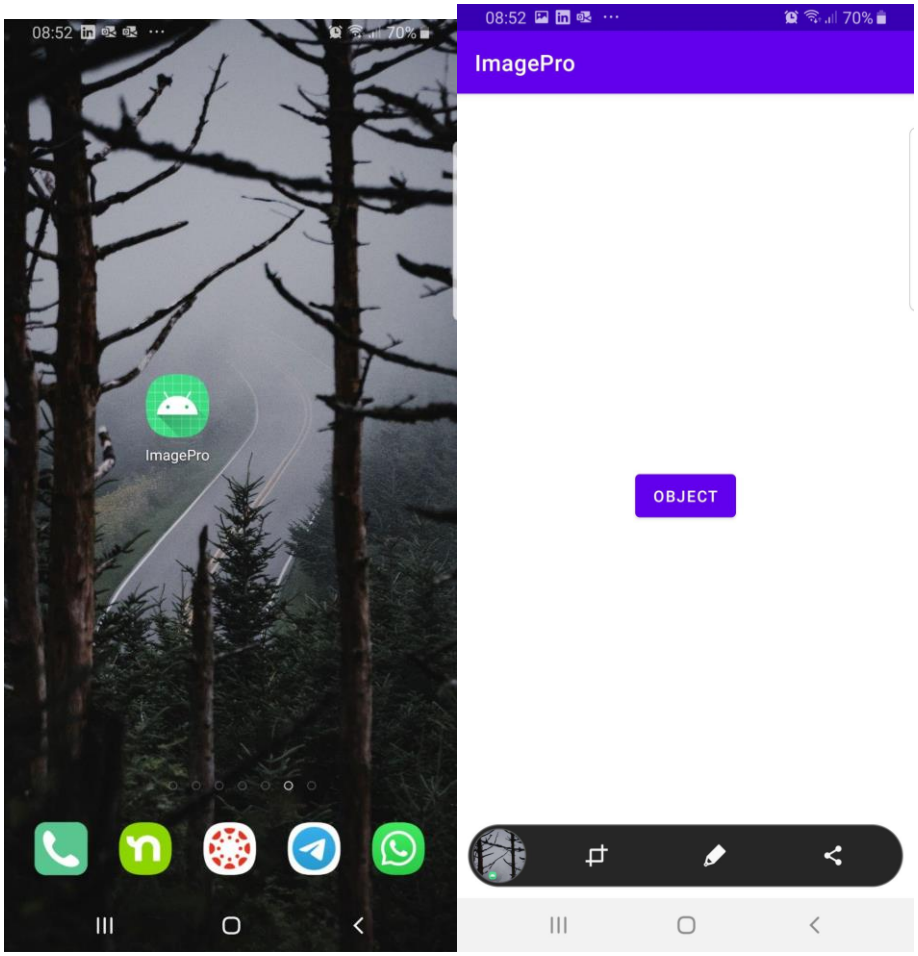
Drag and drop the mobile app.apk onto your phone. Then Go to your file and check the download or recent transfer. Then click on app-debug and download. You might get a message saying ‘unidentified source’.

You would need to allow the application to still be downloaded by the safety prompt message.



Here is the app with the android icon when you press the My Files. Then you will need to select the app file of fer.apk and it will start downloading.

Here is the android application before and after installing the application - click the object button to activate the FER(facial expression recognition).



On the left is the completed application and it should look like this when you open it.

If there an issue using the app follow this part :

You then click the icon object which is shown in the screenshot. Then the app will first close down if it hit a bug. You will need reopen the app and click object button again and it will then work fine or repeat the process a few times and hopefully it works for you.

This is what the end result should look like at the end of the tutorial. Please make sure you are in good lighting as the accuracy goes down when there is a lack of light – it may result in the wrong facial expression output.