



A report submitted in partial fulfilment of the requirements for the degree  
of Bachelor of Science (BSc) in

**COMPUTER SCIENCE**

**UNIVERSITY OF THE WEST OF ENGLAND**

**FACIAL EMOTION RECOGNITION  
FOR MUSIC RECOMMENDATION SYSTEM**

By

YIE NIAN CHU

Supervisor: Craig Duffy

School of Computing  
and Creative Technologies

UNIVERSITY OF THE WEST OF ENGLAND

Date of submission: March 5, 2024  
Word count: WORD COUNT

# DECLARATION

I, Yie Nian Chu confirm that the work presented in this report is my own. Where information has been derived from other sources, I confirm that this has been indicated in the report.

---

Yie Nian Chu

# **ABSTRACT**

## **ACKNOWLEDGEMENTS**

# ACRONYMS

**AI** Artificial Intelligence

**CNNs** Convolutional Neural Networks

**FER** Facial Emotion Recognition

**GD** Gradient Descent

**k-NN** K-Nearest Neighbors

**LR** Linear Regression

**ML** Machine Learning

**MLP** Multi-Layer Perceptron

**MSE** Mean Squared Error

**RF** Random Forest

**SVM** Support Vector Machine

# CONTENTS

<b>Declaration</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>Acknowledgement</b>	<b>3</b>
<b>Acronyms</b>	<b>4</b>
<b>List of Figures</b>	<b>7</b>
<b>1 Introduction</b>	<b>8</b>
1.1 Background .....	8
<b>2 Literature Review</b>	<b>9</b>
2.0.1 Introduction of Music .....	9
2.0.2 Evolution and Diversity of Music .....	10
2.0.3 Music and Emotion.....	12
2.1 Artificial Intelligence and Machine Learning .....	12
2.1.1 Artificial Intelligence.....	12
2.1.2 Machine Learning .....	13
2.1.3 Machine Learning Algorithms .....	13
2.1.3.1 Linear Regression.....	13
2.1.3.2 Support Vector Machine.....	15
2.1.3.3 Random Forest.....	16
2.1.3.4 K-Nearest Neighbors.....	17
2.1.3.5 Neural Networks .....	18
2.1.3.5.1 Multi-Layer Perceptron.....	18
2.1.3.5.2 Convolutional Neural Networks .....	18

2.2	Facial Emotion Recognition .....	20
2.3	Music Recommendation based on FER .....	25
<b>3</b>	<b>Requirements</b>	<b>26</b>
<b>4</b>	<b>Methodology</b>	<b>27</b>
<b>5</b>	<b>Design</b>	<b>28</b>
<b>6</b>	<b>Implementation</b>	<b>29</b>
<b>7</b>	<b>Project Evaluation</b>	<b>30</b>
<b>8</b>	<b>Conclusion and Future Outlook</b>	<b>31</b>
8.1	Conclusion .....	31
8.2	Future Outlook .....	31
	<b>References</b>	<b>32</b>
	<b>Appendix</b>	<b>33</b>
	Appendix A .....	33
	Appendix B .....	34

# LIST OF FIGURES

2.2	Linear Regression .....	14
2.3	Gradient Descent .....	14
2.4	Support Vector Machine .....	15
2.5	Random Forest .....	16
2.6	K-Nearest Neighbors .....	17
2.7	Multi-Layer Perceptron.....	18
2.8	Convolutional Neural Networks .....	19
2.9	Facial Landmarks .....	20
2.13	Linear Regression Classification Accuracies Table .....	23
2.14	Random Forest Classification Accuracy .....	24
2.15	Support Vector Machine Accuracy .....	24



# **1. INTRODUCTION**

## **1.1. BACKGROUND**

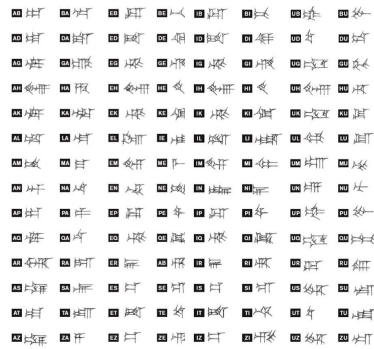
## **2. LITERATURE REVIEW**

### **2.0.1. Introduction of Music**

The development of human communication and social cohesiveness is closely linked to the origins of music. Prothumans probably created the first musical sounds through rhythmic awareness and vocalizations with variable pitch before musical instruments were invented. Primitive humans showed the ability to compose music even before they learned to speak. The development of early human civilizations may have benefited from this musical talent by fostering social cohesion and collaboration.

The comparison of early speech or the pitched sounds of animals to historical music demonstrates how primitive music was as a medium of expression before sophisticated language. A tight relationship exists between the growth of rhythmic awareness and the capacity to produce percussion sounds and the formation of music in the evolutionary timeline. The ability to perceive and produce music was probably important for social interactions when *Homo neanderthalensis* and *Homo sapiens* evolved. One major explanation for the evolution of music is the innate human desire to collaborate and form social contexts for doing so. Early human society appears to have relied heavily on music, judging from its role in fostering social cohesiveness, entertaining, enabling dance, and supporting ritualistic behaviors.

Essentially, the origins of music can be found in the basic human urge for collaboration, socialization, and the development of cultural contexts that support these endeavors. Even in its most primitive incarnations, music influenced human communities and may have helped *Homo sapiens* outperform other hominid species in terms of cultural development.



(a) Cuneiform Alphabet  
Image from ?



(b) World's Earliest Music Composition  
Image from ?

## 2.0.2. Evolution and Diversity of Music

The history of music is extensive and ancient, and it is an essential part of human culture. In Syria, a cuneiform "alphabet" (a) containing the earliest known written composition of music was found (b) ?. It is thought to have been composed some 3400 years ago. The oldest musical instruments, according to archaeological evidence, date back approximately 40,000 years, but the history of music is far older still ?. Despite not yet being documented in the archaeological record, these instruments offer a window into far older musical endeavors. Proto-musical evolution most likely started about 400,000 years ago, based on the social brain hypothesis ?. Musical traditions developed further as people left Africa and spread over the world, reaching a turning point during the Holocene. This historical voyage demonstrates the persistent and varied impact that music has played in human history. Moving forward, there are so many types of music, each representing unique expression of culture, emotion and creativity:

- **Classical Music:** a diverse and evolving tradition, extending beyond the commonly associated period of 1750 to 1820 and encompassing composers from Bach to contemporary artists. It serves as a living and influential force in the world of music, shaping compositions for orchestras, chamber ensembles, solo performers, and even finding unexpected expressions in various genres, from video-game scores to popular music. ?
- **Jazz:** originating in early 20th-century New Orleans, is characterized by complex harmony, syncopated rhythms, and a focus on improvisation. Emerging from a rich tradition of ragtime and blues, jazz evolved into a versatile genre that expanded its influence globally, encompassing popular music standards, modal music, and

even avant-garde compositions.?

- **Blues:** both a musical form and genre, initially associated with melancholy themes, has evolved to encompass a broader range of subjects and emotions, aiming to uplift through music.? Characterized by specific chord progressions, a walking bass, call and response, and unique features like microtonality and flattened 'blue' notes, blues is known for its distinct sound and expressive style. ?
- **Electronic:** utilizes diverse sound sources, ranging from recorded sounds captured by microphones to those generated by electronic oscillators and complex computer installations. Typically played back through loudspeakers, electronic music can be created using various technologies, with the exception of "live electronic music," which involves real-time performance. ?
- **Country:** originating in rural Southern and Western areas in the early 20th century, was initially labeled "hillbilly music" before adopting the official term "country and western music" in 1949. Rooted in the ballads, folk songs, and popular tunes of English, Scots, and Irish settlers, country music gained commercial recognition in the early 1920s, marked by its realistic portrayal of rural life contrasting with the sentimental tone prevalent in popular music of that era. ?
- **Hip-Hop:** a cultural movement that gained widespread popularity in the 1980s and '90s, serves as the foundational music for rap—a style featuring rhythmic and/or rhyming speech, which became the movement's enduring and influential art form. Hip-hop, beyond its musical dimension, encompasses diverse elements such as graffiti art, breakdancing, and social activism, reflecting a multifaceted expression of urban culture and creativity. ?
- **Rock:** a form of popular music that emerged in the 1950s and by the end of the 20th century became the dominant global music genre, influencing the recording industry, international retail, and radio and television playlists. While dictionary definitions often focus on its strong beat and instrumentation, the cultural significance of rock lies in its social and ideological distinctions from other music genres, particularly its development as a term to distinguish certain attitudes and practices from those associated with pop music.?

### **2.0.3. Music and Emotion**

Since ancient times, people have been fascinated by the paradoxical connection that exists between music and emotion. Even though music is an abstract art form that appears to be removed from everyday life, it has a profound potential to evoke strong emotional responses. This ability of music to trigger strong emotions is also demonstrated in other social circumstances, such as advertising. This encounter is further enhanced by the relationship that exists between music and our individual life experiences. Emotions, influenced by these encounters, provide our perception and thought processes a personalized meaning that connects the abstract quality of music to the concrete events of our everyday life. This complex tapestry that highlights the profound influence of music on our emotional environment is created by the blending of music, emotion, and human experience ?.

Numerous musical elements that have been thoroughly explored are combined to convey emotions through music. Pace, mode, harmony, interval, rhythm, sound level, timbre, timing, articulation, accents, tone attacks and decays, and vibrato are some of these characteristics. Emotion in music is expressed through compositional elements as well as performance elements. Still, it's not an easy task to express feelings through music. Certain musical elements can be employed to convey a range of moods, showing that certain elements are not always reliable predictors of a certain emotion. The Lens Model ?, which characterizes emotional expression in music as involving probabilistic and partially redundant auditory cues, sheds more light on this complexity. Listeners combine several cues for successful emotion recognition, and the redundancy of cues allows for a high level of emotion recognition through different combinations, offering room for creativity and personal expression ?.

## **2.1. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

### **2.1.1. Artificial Intelligence**

Artificial Intelligence (AI) is a broad field that includes using technology to build machines and computers that can replicate cognitive abilities connected to human intellect. These abilities include making recommendations, language understanding, data processing, and visual perception. AI should be viewed as a group of technologies

incorporated into systems rather than as a stand-alone system that can understand, learn, and respond to complex problems.

## 2.1.2. Machine Learning

Machine Learning (ML) is a branch of AI focused on enabling machines and systems to learn and enhance their performance through experience. Instead of relying on explicit programming, machine learning employs algorithms to analyze vast datasets, derive insights, and subsequently make informed decisions. These algorithms continually improve their performance as they are exposed to more data. The outcomes of this learning process are the machine learning models, which become more proficient with increased exposure to data.

## 2.1.3. Machine Learning Algorithms

### 2.1.3.1. Linear Regression

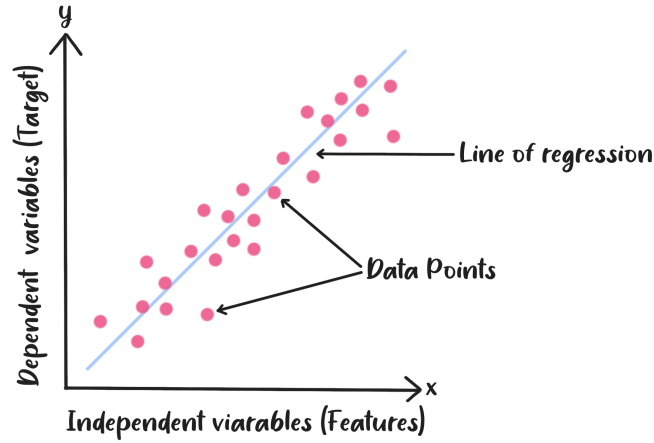
Linear Regression (LR) is a supervised learning algorithm used to model the relationship between a dependent variable (target) and one or more independent variables (features). The fundamental assumption of LR is that there exists a linear relationship between the input variables and output. ?

$$y = mx + b \quad (2.1)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \quad (2.2)$$

A LR model can be represented by the equation 2.2 where  $Y$  represented the dependent variable.  $\beta_0$  is the y-intercept,  $\beta_1, \beta_2, \dots, \beta_n$  are the coefficients and the  $\epsilon$  is an error term, representing the unobserved factors that affect  $Y$  but are not accounted for by the model. The logic of it is same as Linear Equation (2.1) where using the gradient of the line ( $m$ ), the value of  $x$  and the y-intercept ( $b$ ) to get the value of  $y$ , which is what we are trying to predict.

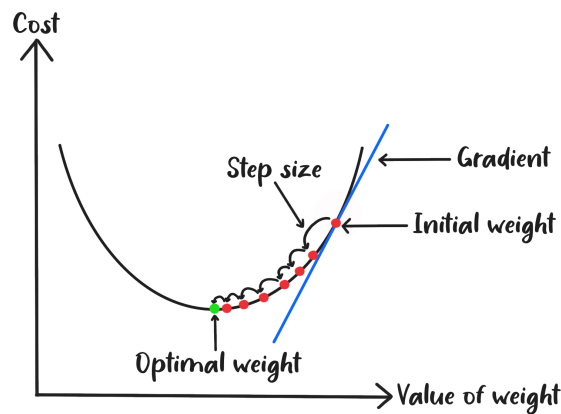
While using LR, the main objective is to find the values of  $\beta_0, \beta_1, \dots, \beta_n$  that minimize the error between the predicted value ( $\hat{Y}$ ) and the actual value ( $Y$ ) in the training data.



**Figure 2.2:** Linear Regression

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.3)$$

Therefore, Mean Squared Error (MSE), a cost function to measures the average squared difference between  $\hat{Y}$  and  $Y$ , is introduced to quantify the goodness of fit of the model to the training data. If the current result is not optimized, Gradient Descent (GD), an optimization algorithm, will be used to adjust the coefficients,  $\beta_0, \beta_1, \dots, \beta_2$ , towards the direction that minimizes the MSE until it reach the convergence (Figure 2.3).



**Figure 2.3:** Gradient Descent

### 2.1.3.2. Support Vector Machine

Support Vector Machine (SVM) is a supervised ML algorithm where we used for classification, regression and outliers detection. The main goal of it is to find the hyperplane that best separates the data into different classes based on statistical approaches. ?

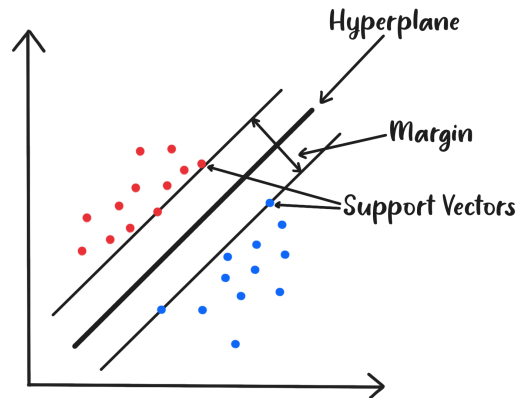


Figure 2.4: Support Vector Machine

While using SVM, the greater the margin, the better the result would be as it has better generalization to new or unseen data. There are two types of SVM for classification, which are Linear SVM and Non-linear SVM. A linear SVM finds the optimal hyperplane that maximizes the margin between classes for linearly separable data as Figure 2.4.

$$f(x) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \quad (2.4)$$

The decision function of linear SVM (Equation 2.4) is used to defines the ability to classify data points into different classes. When the result is greater than or equal to zero, the prediction would be positive. If  $f(x)$  is less than zero, the decision function predicts the negative class.

For non-linearly separable data, SVM uses kernel functions such as polynomial, sigmoid and radial basis function (RBF), to map the data into a higher-dimensional space where hyperplane can separate the classes. The equation 2.5 is the decision



$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b\right) \quad (2.5)$$

$$K(x, x_i) = (x \cdot x_i + c)^d \quad (2.6)$$

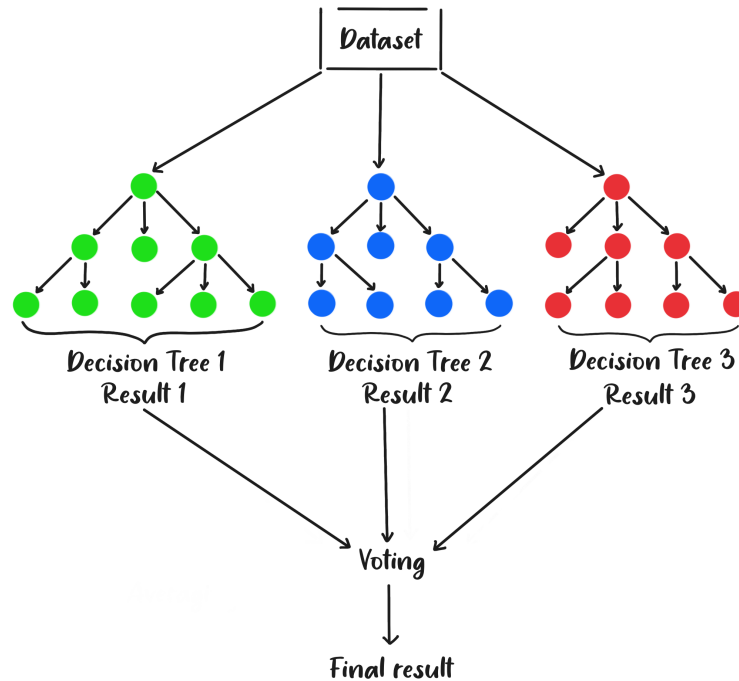
$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (2.7)$$

$$K(x, x_i) = \tanh(\alpha x \cdot x_i + c) \quad (2.8)$$

function of non-linear SVM. The kernel function, denoted by  $K(x, x_i)$  in the equation, would be replaced by equation 2.6 if a polynomial kernel were used. Similar with the other kernels, if the RBF kernel is employed, it would be exchanged with equation 2.7, and for sigmoid kernel, it would be replaced with equation 2.8.

### 2.1.3.3. Random Forest

Random Forest (RF) is an ensemble learning algorithm that belongs to the family of decision tree-based methods. A group of decision trees that have been trained on various dataset subsets make up the forest of RF, and the final result is derived from averaging the predictions of each individual tree. ?



**Figure 2.5:** Random Forest

As Figure 2.5 shown, RF builds multiple Decision Trees and combines them to get a

more accurate and stable prediction than any individual model. This process is called bagging which is one of a type of ensemble learning. In the process, the entire dataset is separated into subsets and each decision tree is trained individually on a subset that is selected at random. This adds variety and unpredictability to the trees. The training process will then generate results for each model, and the final output is determined by the "votes" for a class from each tree. The class with the majority of votes is chosen as the final prediction.

#### 2.1.3.4. K-Nearest Neighbors

K-Nearest Neighbors (k-NN) is a intuitive supervised machine learning method used for both classification and regression tasks. The main idea of k-NN is to predict the label of new data point based on its k-nearest data points in the feature space. ?

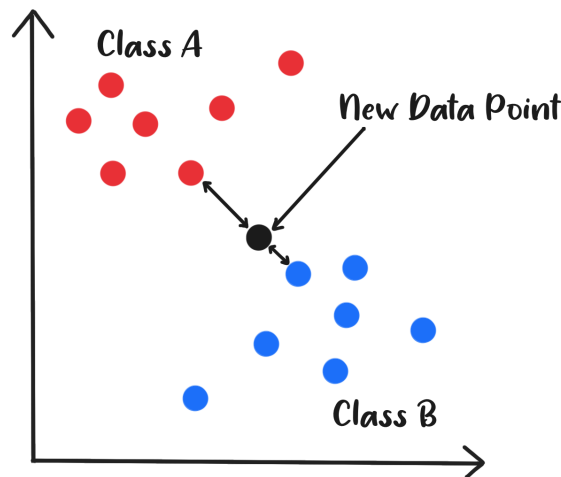


Figure 2.6: K-Nearest Neighbors

k-NN uses distance metric (Equation 2.9) to calculate how similar two data points are to one another. k-NN finds the k training data points that, according to the selected distance metric, are closest to a given data point. In classification tasks, the majority class among a new data point's k-nearest neighbors predicts the class that the data point will fall into.

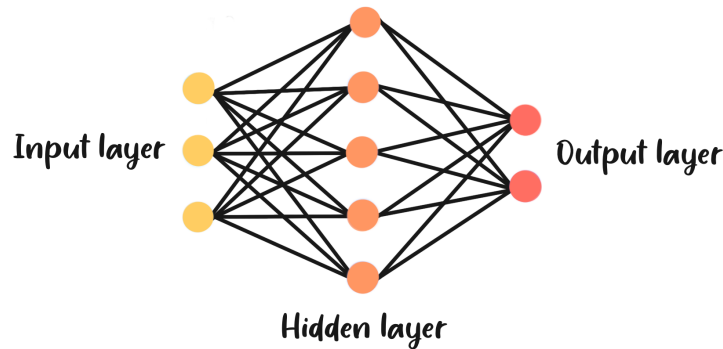
$$d(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.9)$$

$$\hat{Y} = \operatorname{argmax}_y \left( \sum_{i=1}^k I(y_i = y) \right) \quad (2.10)$$

The key hyperparameter of k-NN is the value of  $k$  (Equation 2.10), representing the number of nearest neighbors to consider. The choice of  $k$  can significantly impact the performance of the algorithm, and it is often selected through cross-validation.

### 2.1.3.5. Neural Networks

#### 2.1.3.5.1 Multi-Layer Perceptron



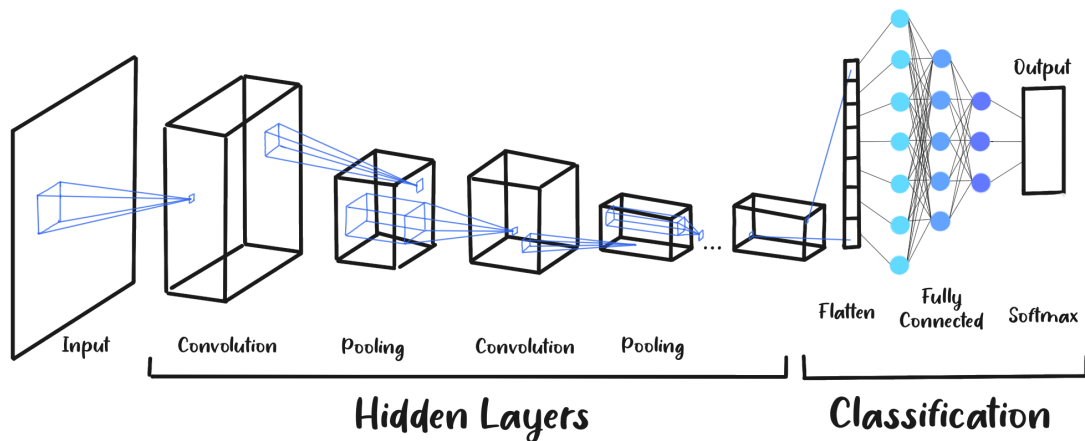
**Figure 2.7:** Multi-Layer Perceptron

An input layer, one or more hidden layers, and an output layer are the minimum number of nodes that make up a Multi-Layer Perceptron (MLP) feedforward artificial neural network type (Figure 2.7). All nodes in these layers—aside from those in the input layer—are linked using a specific weight and employ a nonlinear activation function. Because of its nonlinearity, the network may learn and carry out more complicated tasks as well as represent intricate connections between the input and output. A MLP model is trained by comparing its output to the expected output and propagating errors back through the network to alter the weights. This process is known as backpropagation.

?

#### 2.1.3.5.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a family of deep learning algorithms that are mostly utilized for processing input that has a grid structure, like images. ?



**Figure 2.8:** Convolutional Neural Networks

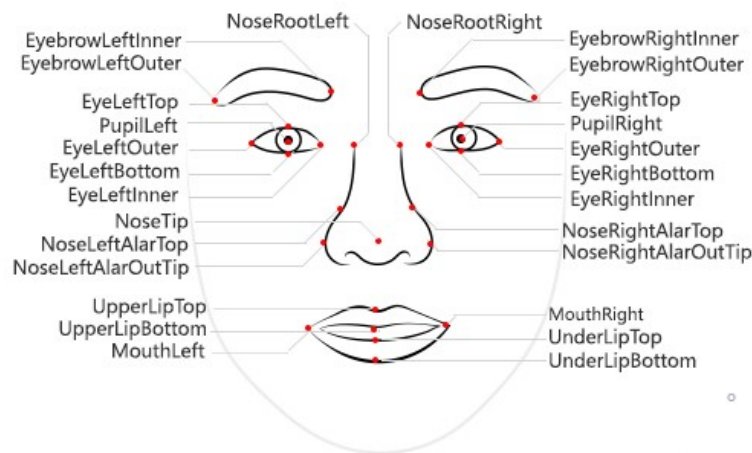
CNNs are designed to adaptively learn spatial hierarchies of features from the data. This learning process includes convolution layers, pooling layers, flatten layer, and fully connected layers. By applying filters, also known as kernels, to the input, these layers carry out the convolution process and produce feature maps. Local elements like textures and edges are captured throughout this procedure. Pooling layers, which come after convolutional layers, help to reduce the number of parameters and computation in the network by reducing the spatial dimensions (width and height) of the input volume.

The features from the input image are retrieved by the convolutional and pooling layers, and the next stage is to categorize the features which is done in flatten layer. The feature maps are converted into a one-dimensional vector in the flatten layer, which is necessary for fully connected layers. The flattened vector is then fed into the fully connected layers (which resemble the standard neural network layers with fully connected nodes) for the classification task. These fully connected layers divide the image into discrete groups based on the high-level characteristics found in the preceding levels.

The last layer of layers in the network architecture play the crucial job of generating the output. The output layer typically uses a softmax activation function in multi-class classification settings to translate the network's raw output into probabilities given to each class. The output node with the highest probability is then chosen to determine the anticipated class.

## 2.2. FACIAL EMOTION RECOGNITION

Human emotions can be inferred from facial expressions. Deciphering these signs of emotion has become a popular research topic in the fields of Human Computer Interaction and Psychology. ? The development of Facial Emotion Recognition (FER) technology has been significantly aided by technological advancements, particularly with the introduction of ML and Pattern Recognition.



**Figure 2.9:** Facial Landmarks  
Image from ?

FER is a broad field that intersects with Computer Science, AI, Psychology and other fields. It involves analyzing a person's facial expressions in still images and videos in order to determine their emotional state. A three-step approach is used in the methodology: face detection, facial expression identification, and categorization of the expression into a certain emotional state. Facial landmark (Figure 2.9) detection and analysis of changes in their positions are key components of this complex process. FER attempts to offer insights into people's emotional experiences by identifying muscular contractions linked to various emotions from visual clues found in facial expressions.

As stated by Tarnowski et al. (2017), the creative feature extraction from facial expressions using coefficients that detailed aspects of emotional states is what makes the research successful. They distinguished between seven different emotional states: happiness, sorrow, surprise, wrath, fear, and contempt, as well as the more subdued

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	881	125	2	155	340	48	101
joy	106	922	7	238	115	1	154
surprise	13	1	1135	8	8	390	13
anger	130	151	6	862	81	2	120
sadness	229	130	33	101	823	104	88
fear	41	0	220	5	88	871	4
disgust	76	147	73	107	21	60	996

(a) Confusion Matrix for k-NN Classifier

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	1160	75	9	29	644	61	41
joy	81	1178	0	57	141	0	129
surprise	3	0	1153	4	2	426	10
anger	58	137	0	1346	44	0	88
sadness	122	13	5	0	561	75	2
fear	8	1	308	1	74	910	2
disgust	44	72	1	39	10	4	1204

(b) Confusion Matrix for MLP Classifier

Tables from ?

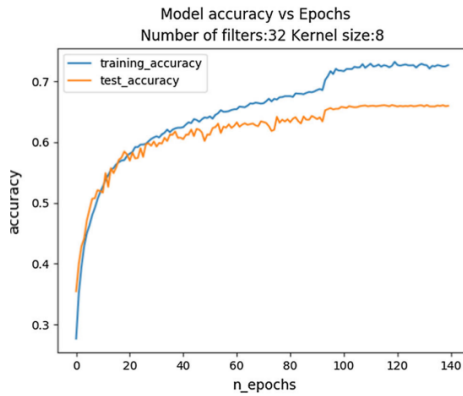
displays of neutrality. For feature computation, they employed a three-dimensional facial model as an alternative to conventional two-dimensional approaches. This enables them to collect more detailed and subtle data, which may improve the accuracy of identifying emotions. On top of that, they used the MLP neural network and the k-NN classifier to classify each emotional state. According to their research and comparative analysis, the MLP is more accurate than the k-NN in classifying emotional states, with a 73% classification accuracy compare to a 63% accuracy for k-NN ?.

Mellouk et al. (2020) showed in their study that FER may be achieved with great accuracy and effectiveness by utilizing deep learning techniques. They provided a thorough analysis of multiple FER databases, emphasizing their diversity in terms of picture, video content, lightning circumstances, and demographic variances—all of which are important determinants of FER performance—in order to guarantee the credibly of the results.

Traditional facial recognition techniques included manually defining and extracting features from facial photos, a procedure that was frequently less flexible and efficient. Examples of these techniques include Local Binary Patterns (LBP), Facial Action Coding System (FACS), Local Directional Patterns (LDA), and Gabor wavelet. CNNs and LSTMs can automatically extract and learn complex patterns from facial data, according to Mellouk et al., which will improve the reliability and accuracy of emotion recognition. Furthermore, the difficulties that traditional approaches faced—such as variances in facial characteristics due to diverse demographics, occlusions, and data diversity—are resolved with the use of deep learning, increasing their versatility and effectiveness. The study also examines preprocessing methods including image scaling, cropping, normalization, and data augmentation that are crucial for improving the accuracy of these deep learning models.

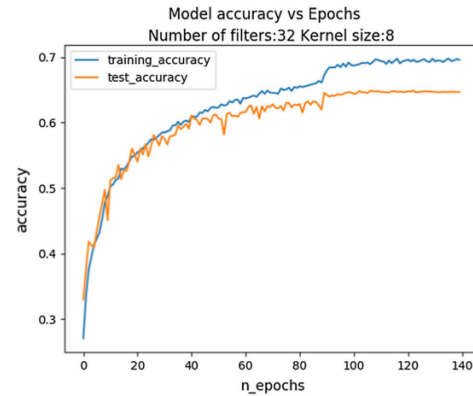
With the help of deep learning and the preprocessing techniques they found, the

outcome demonstrates proficiency in accurately classifying the fundamental emotions, with some models reaching over 90% accuracy under specific circumstances ?. It indicates that as machines improve at deciphering human emotions, interactions between humans and machines may become more intuitive and natural.



(a) Accuracy variation: Model1

(a) Accuracy variation with 32 Filters and 8 kernel size for Model 1



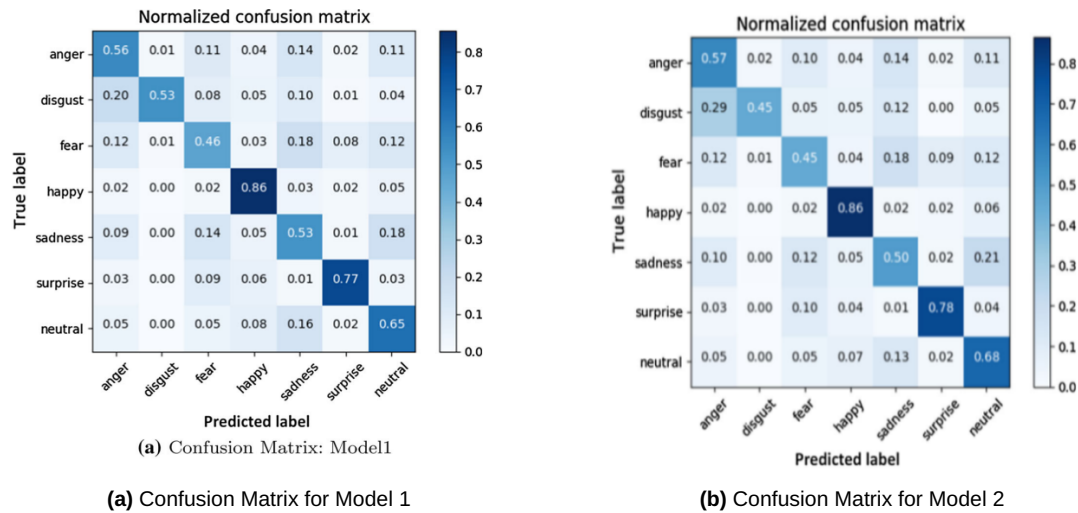
(b) Accuracy variation: Model2

(b) Accuracy variation with 32 Filters and 8 kernel size for Model 2

Graphs from ?

Based on the findings of Agrawal et al.'s work, the kernel size and the number of filters significantly impact CNNs accuracy. Using the FER-2013 dataset as their primary emphasis, two CNN architectures is put forth after a thorough analysis of various kernel sizes and filter counts. To find the optimal set of parameters that could yield the best convergence and accuracy, Agrawal et al. ran tests with the combination of 6 different kernel sizes (2, 4, 8, ..., 64) and 8 different number of filters (2, 4, 8, ..., 256). They discovered that when network depth increased, a network with 32 filters and an 8 kernel size demonstrated a discernible gain in accuracy (Figure 2.11a, Figure 2.11b).

Even while Model 2 is simpler due to its constant kernel size, lack of dropout layers, and fully connected layers, it was nevertheless able to achieve an accuracy of 65% on the FER-2013 dataset, which is comparable to human performance ?. Furthermore, in comparison with other emotions, the proposed models were able to categorize happiness and surprise with a higher degree of accuracy, which is consistent with people's challenges in picking out distinct emotions (Figure 2.12a, Figure 2.12b).



Graphs from ?

Gestures	Recognition Accuracy
Neutral	99.00%
Smile	98.50%
Anger	98.50%
Scream	99.50%
Overall	98.88%

**Figure 2.13:** Linear Regression Classification Accuracies Table  
Table from ?

A pivotal study by Naseem et al. (2010) incorporates the analysis of facial expressions, recognizing them as crucial variations in appearance induced by internal emotions or social communications. In order to evaluate their LR Classification approach, they therefore took into account occlusion modes, brightness changes, and expressions such as scream, smile, rage, and neutral. Notably, the LR Classification algorithm showed an excellent recognition accuracy for all facial expressions tested, averaging 98.88% in a 100D feature space (Figure 2.13). For the screaming expression, the algorithm outperformed other accuracies by achieving an accuracy of 99.5% (Figure 2.13). This great accuracy demonstrates the reliability and efficacy of the LRC approach in handling a wide range of facial emotions.

According to Munasinghe (2018), RF Classifier are capable of handling facial expression variability well and without overfitting. Also, the researcher asserts that facial landmarks (Figure 2.9) provide an accurate feature extraction capability that capture subtle changes in facial emotions. A facial feature vector obtained from these landmarks and normalized to reduce variance in face size is used to discern emotions



	Facial Expression				Total
	<i>Anger</i>	<i>Happiness</i>	<i>Sadness</i>	<i>Surprise</i>	
Training (70%)	31	48	19	58	156
Testing (30%)	14	21	9	25	69
Total	45	69	28	83	225
Success	79%	95%	89%	96%	90%

**Figure 2.14:** Random Forest Classification Accuracy  
Table from ?

with a RF Classifier. With the aid of feature vector, the RF Classifier achieved an average success rate of 90% in classifying four different emotions: anger, happiness, sadness, and surprise (Figure 2.14).

	happy	surprise	fear	angry	sad	disguise
h	144	7	5	10	11	23
s	21	147	6	10	1	15
f	1	5	143	8	19	24
a	11	2	21	132	26	8
s	6	14	18	27	119	16
d	12	13	24	21	6	124

**Figure 2.15:** Support Vector Machine Accuracy  
Table from ?

Li Xia (2014) presents a unique method of facial emotion detection that employs multi-classification SVM. The study proposes a two-on-two classification method which is an innovative approach to overcome the limitations of traditional classification methods like one-against-one (classifier is trained for each pair of classes) and one-against-the-rest (classifier is trained against all other classes combined). With this novel method, the classification process is faster with fewer sub-classifiers and reduced classification errors. The results of this study were impressive, showing the classifier in this investigation demonstrated a high average recognition rate of 92.7% when six distinct emotions were considered, including happiness, surprise, anger, fear, disgust, and sadness.

## 2.3. MUSIC RECOMMENDATION BASED ON FER

From Chakrapani et al.'s approach, music recommendation system with deep learning algorithm could enhance the listening experience by accurately detect and interpret the user's emotions. This is achieved by using CNNs to analyze the user's age, gender, and facial emotion. Based on these data, the system would cater to the user's preferences and present mood. To ascertain the user's emotional state, they used the webcam to take pictures of the user and then processed the image using the CNNs models. The system then provided tailored music recommendations based on the predictions generated by the CNNs models. This approach offers a creative and user-centric alternative for music selection based on emotional cues while streamlining playlist construction and management. ?

Additionally, Athavle et al. discovered that using CNNs model helps a music recommendation system to accurately detect emotions and subsequently recommend music that aligns with the user's mood. They train a CNNs model for emotion detection in their work. While maintaining great precision, this method lowers total system costs and computing time. The system uses real-time emotion detection to work, and then sends the data to the CNNs model to classify the user's emotions. An appropriate playlist will be recommended as soon as the technology determines the user's current feeling, making the user experience engaging and responsive. In order to guarantee optimal classification accuracy and efficacy, they employed categorical cross-entropy as a loss function to manage missing and anomalous values inside the FER2013 dataset. Despite their result being less accurate than Chakrapani et al.'s work (71%), it nevertheless shows that the model is effective and trustworthy in identifying emotions from facial expressions. ?

### **3. REQUIREMENTS**

## **4. METHODOLOGY**

## **5. DESIGN**

## **6. IMPLEMENTATION**

## **7. PROJECT EVALUATION**

## **8. CONCLUSION AND FUTURE OUTLOOK**

### **8.1. CONCLUSION**

### **8.2. FUTURE OUTLOOK**



# BIBLIOGRAPHY

- , G. (2019), 'Descending into ml: Training and loss | machine learning crash course'.  
**URL:** <https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss>
- , M. (2021), 'Country music guide: History and sounds of country music'.  
**URL:** <https://www.masterclass.com/articles/country-music-guide>
- , s.-l. (n.d.), '1.4. support vector machines'.

# APPENDIX

## APPENDIX A

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## APPENDIX B

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.