**A PROJECT REPORT**

**ON**

**MUSIC PLAYER BASED ON EMOTION RECOGNITION**

**IS SUBMITTED TO**

**INSTITUTE OF ENGINEERING AND MANAGEMENT**

SUBMITTEB BY

**RAVI RANJAN**

**Roll no:- 128**

Under the guidance of

**Mr. Swagtam Basu**





**CERTIFICATE OF APPROVAL**

This is the certify that the work embodied in this project entitled MUSICAL PLAYER BASED ON EMOTION RECOGNITION submitted by RAVI RANJAN to the Department of Information Technology, have been carried out under my direct supervision and guidance. The project work has been prepared as per the regulation of Institute of Engineering and Management and I strongly recommended that this project work be accepted in fulfilment of the INOVATIVE project.

**Supervised Under**

**Project Guide Head of Dept.**

Prof. Swagtam Basu Prof. Moutushi Singh

**Abstract**

This project an Emotion based music player is a approach that helps the user to automatically play song based on the emotion of the user and solve the issue of the manual finding od songs for suited mood with high accuracy of Emotion recognition. It recognizes the facial emotion of the user and plays the songs according to the emotion. The emotion are recognized using a Deep Learning Convolutional Neural Network and Haar Cascade Classifier. Then a playlist will be generated through the system according to the mood predicted by the model. The model used for predicting the mood on the training data from FER2013 Dataset.

1. **INTRODUCTION:-**

Music plays a very important role in enhancing an individual’s life as its important medium for relaxing and stress release. In today’s world, many music players are available with different features like fast forward, reverse, various playback speed, local playback, different genre groups etc. these features satisfy the user’s basic requirements. Yet the user has manually select the playlist of songs and select songs based on his moods.

**1.1 Objective:-**

My basic approach is to create Emotion based music player that helps user to automatically play songs based on the emotion of the user. It recognizes the facial emotion of the user and plays the songs according to their emotion. The emotion are recognized using a Convolutional Neural Network and harr cascade frontal face classifier algorithm. The human face is an important organ of an individual’s body and it’s important role for showing behaviour and expression of the user. In this project, the webcam captures the image of the user. Then it will extract the human face feature of the user from captures image. After that extract the facial features of the user from captures image. Facial expression categorized into seven different moods (Happy, Sad, Disgust, Fear, Neutral, Angry, Surprise).

**1.2 Scope:-**

Facial Expression are the great indicator of the showing the state of mind for a person. Indeed, the most natural ways to express emotion is through facial expressions. Humans links the music lyrics from which music they listen and how they feel. It would be helpful if the music player was “smart enough” to sort out the music based on the current state of emotion the person is feeling. This project helps them to create a playlist based on user current emotion of user.

**1.3 Application:-**

* Automatically play song based on the emotion of the user.
* Smart TV
* Recommendation on YouTube.
* Can be used for Blind people.
* Personal Assistant

1. **Machine Learning and Deep Learning:-**

Machine Learning is a field of computer science that uses statistical techniques to give computer systems the ability to “learn”(i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

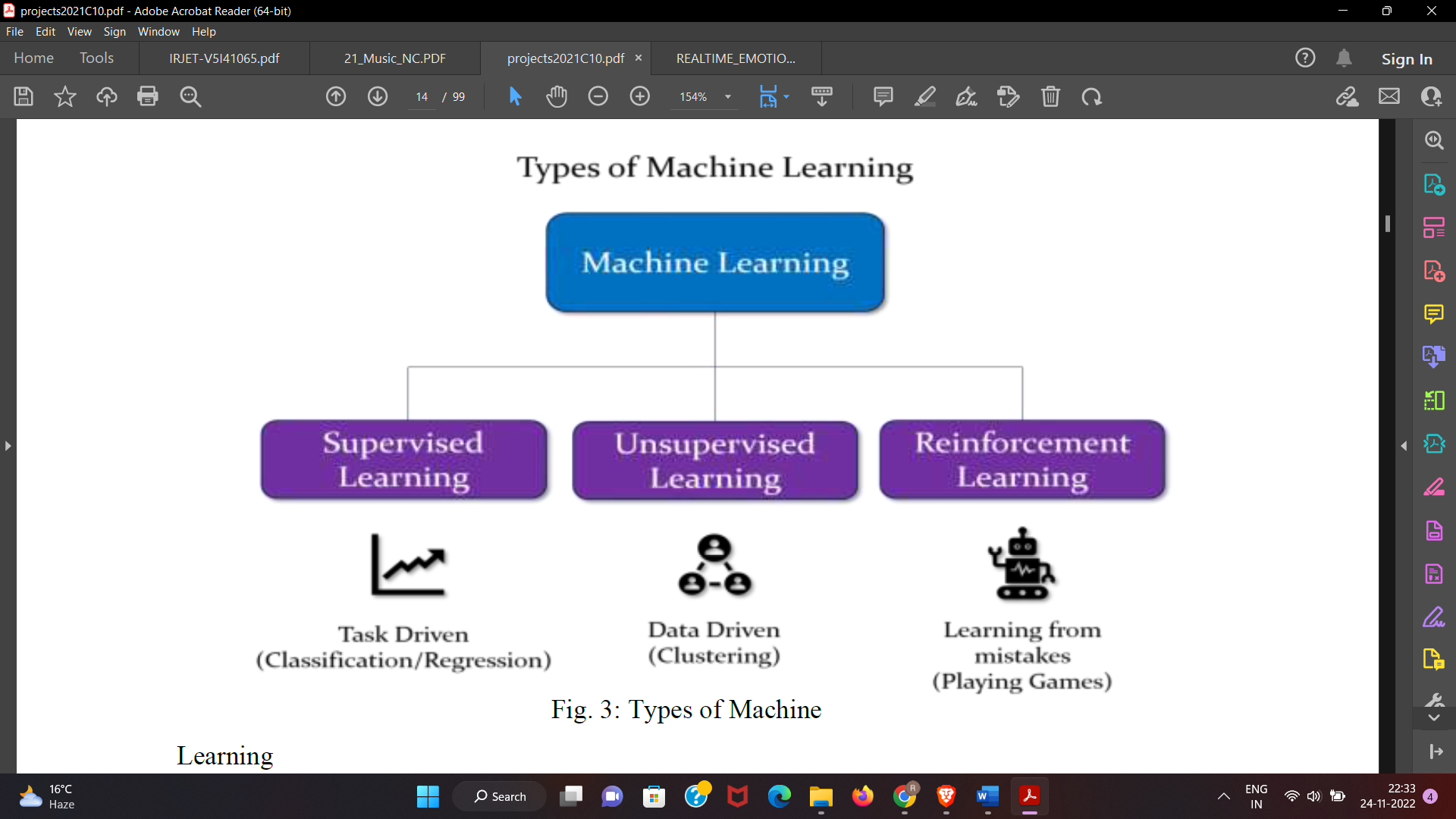
study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs.

Machine Learning Algorithms can be classified into 3 types as follows:

1. Supervised learning

2. Unsupervised Learning

3. Reinforcement Learning

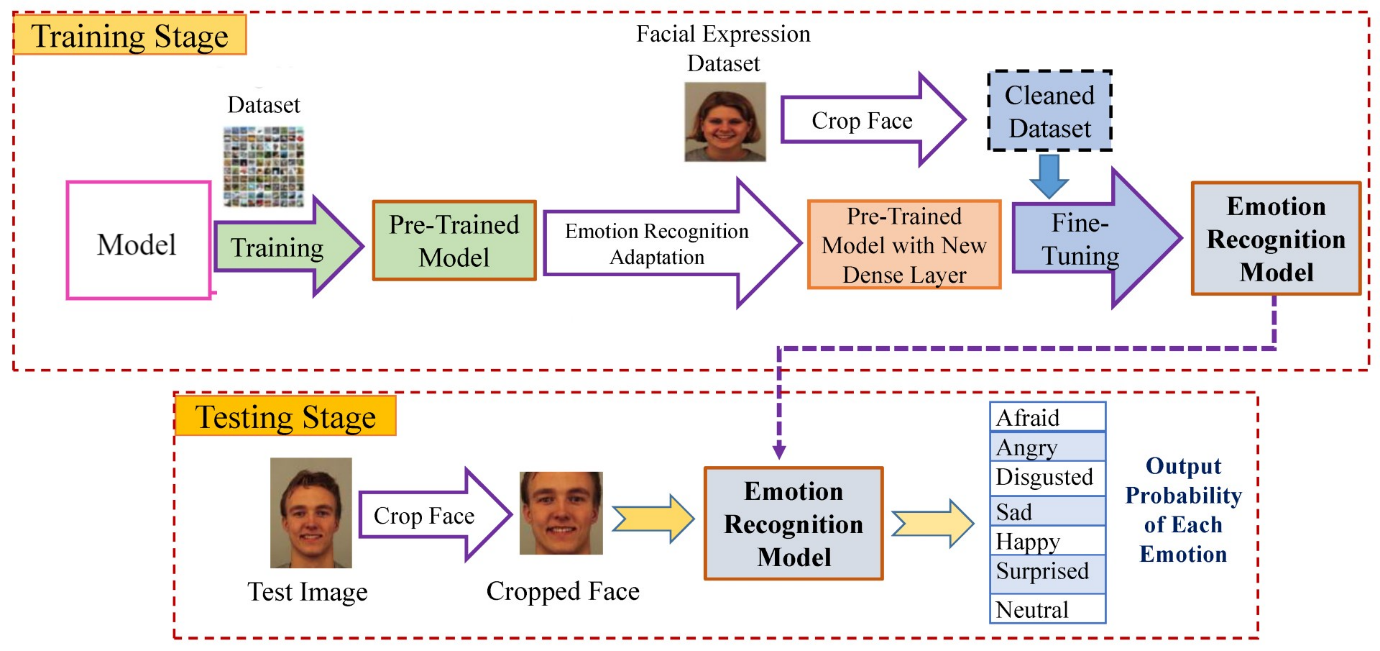


Convolutional Neural Networks(CNN) is a specific type of Artificial Neural Network which are widely used for image classification.

CNN is a type of deep Learning model for processing data that has a grid pattern and designed to automatically and adaptively learn spatial hierarchies of features, from low-to high-level patterns. CNN is a mathematical construct that is typically composed with three types of layers:- convolution, pooling and fully connected layers. The first two (convolutional and pooling) perform feature extraction, whereas third (a fully connected layer) maps the extracted features into final output, such as classification in our project. This project consists of 4 module – faces detection, feature extraction, emotion detection and songs classification.

A Convolution layer (third layer) plays a key role is CNN, which is composed of the stack of mathematical operation specialized in linear operation. In digital images, pixel values are stored in a two dimensional (2D) grid, i.e., an array of numbers and a small grid of parameters called kernel, an optimizable feature extractor, is applied at each image position, which makes CNNs highly efficient for image processing, since a feature may occur anywhere in the image. As one layer feeds its output into the next layer, extracted features can hierarchically and progressively become more complex. The process of optimizing parameters such as kernels is called training, which is performed so as to minimize the difference between outputs and ground truth labels through an optimization algorithm called backpropagation and gradient descent, among others.

1. **METHODOLOGY**

****

**3.1 Face Detection:**

**The Viola-Jones Algorithm** is an object-recognition framework that allows the detection of image features in real-time. Viola-Jones is quite powerful and its application has proven to be exceptionally notable in real-time face detection. The framework is still a leading player in face detection alongside many of its CNNs counter parts. The Viola-Jones Object Detection Framework combines the concepts of **Haar-like Features, Integral Images, the AdaBoost Algorithm, and the Cascade Classifier** to create a system for object detection that is fast and accurate. Viola-Jones outlines a box and searches for a face within the box.

Before detecting a face, the image is converted into grayscale, since it is easier to work with and there’s lesser data to process. The Viola-Jones algorithm first detects the face on the grayscale image and then finds the location on the colored image.

**3.2 Facial Feature Extraction:**

Convolution neural network (CNN) is an efficient recognition algorithm which is widely used in pattern recognition and image processing. It has many features such as simple structure, less training parameters and adaptability. CNN is a class of **deep learning neural networks**. CNNs represent a huge breakthrough in image recognition. They’re most commonly used to analyze visual imagery and are frequently working behind the scenes in image classification. Image classification is the process of taking an **input** (like a picture) and outputting a **class** or a **probability** that the input is a particular class (“there’s a 90% probability that this input is an image”).This is important when we are to design an architecture which is not only good at learning features but also is scalable to massive datasets. A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

**3.4 Emotion Detection:**

This module deals with the usage of the convolutional neural network which uses a MobileNet architecture and The Viola-Jones Algorithm which is used to detect the presence of face and extract the features needed form a pattern that matches the facial expression. The CNN is trained manually for the classification of 7 emotion states (happy, sad, surprise, angry, disgust, fear, neutral)

The core layer of MobileNet is depth wise separable filters, named as Depth wise Separable Convolution. The network structure is another factor to boost the performance. Finally, the width and resolution can be tuned to the trade off between latency and accuracy. Depth wise separable convolutions which are a form of factorized convolutions which factorize a standard convolution into a depth wise convolution and a 1×11×1 convolution called a pointwise convolution. In MobileNet, the depth wise convolution applies a single filter to each input channel. The pointwise convolution then applies a 1×11×1 convolution to combine the outputs the depth wise convolution

n.

**3.5 Song Classification:**

This module deals with the procedure that was used to identify the mapping of each song with its mood. Based on these features, we need to train an artificial neural network which successfully classifies the songs in 7 classes with a higher accuracy . The songs have to manually labelled. All the features will extracted using Python and relevant packages

**3.6 Dataset :-**

We are using Fer-2013 Dataset in this project. The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less cantered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

**3.7 System Requirements**

Operating system: windows, Mac OS, Linux

Libraries: OpenCV, NumPy, Keras

**3.8 How it will Work:-**

* The computer vision captures the image of the user to consider it to identify the facial expression.
* The matplotlib will plot the possible libraries for the desired output.
* The CNN Haar Cascade analysis is carried out on the image and predictions are obtained.
* Then emotion is identified from the predictions and the rectangle bounding of the detected face is done.
* Based on the emotion represented by the facial expression opens the specific music file and plays music



1. **FUTURE SCOPE**

We can use this project as :-

* Automatically play song based on the emotion of the user using Spotify API
* Smart TV that choose channels or movies web series according to our mood.
* Recommendation on YouTube.
* Can be used for Blind people which is more convenient for them. They can show their emotion and different application works according to their mood.
* Personal Assistant based of Emotion Recognition

1. **CONCULUSION**

In this project, we are generating the playlist according the emotion of the user, we developed an application for predicting the emotion of the user using Convolution neural networks and for generating the playlist. The model classifies 7 different facial emotions from the image dataset. Our work aims to achieve the highest possible accuracy while not compromising the real-time aspect to apply to the real-world scenario.

1. **REFRENCES**

* <https://ijariie.com/AdminUploadPdf/REALTIME_EMOTION_BASED_MUSIC_PLAYER_USING_IMAGE_PROCESSING_ijariie17839.pdf>
* <https://www.irjet.net/archives/V5/i4/IRJET-V5I41065.pdf>
* <http://www.ijirset.com/upload/2020/june/21_Music_NC.PDF>
* <https://www.mygreatlearning.com/blog/viola-jones-algorithm/#sh3>
* <https://medium.com/analytics-vidhya/image-classification-with-mobilenet-cc6fbb2cd470>