ABSTRACT

Music plays a very important role in human's daily life. Everyone wants to listen music of their individual taste, mostly based on their mood. Users always face the task of manually browsing the music and to create a playlist based on their current mood.

The proposed project is very efficient which generates a music playlist based on the current mood of users. Facial expressions are the best way of expressing ongoing mood of the person. The objective of this project is to suggest songs for users based on their mood by capturing facial expressions. Facial expressions are captured through webcam and such expressions are fed into learning algorithm which gives most probable emotion. Once the emotion is recognized, the system suggests a play-list for that emotion, thus saves a lot of time for a user.

Once the emotion is detected by CNN then the emotion is used by Spotify API and then the Spotify API generates a playlist according the emotion of the user.

Keywords*:* Face detection, Emotion recognition, Webcam, CNN classification, Spotify API, Music Playlist.

LIST OF FIGURES

TABLE OF CONTENTS

1. INTRODUCTION

Music plays an important role in our daily life. Users have to face the task of manually browsing the music.

Computer vision is a field of study which encompasses on how computer see and understand digital images and videos.

Computer vision involves seeing or sensing a visual stimulus, make sense of what it has seen and also extract complex information that could be used for other machine learning activities.

We will implement our use case using the Haar Cascade classifier. Haar Cascade classifier is an effective object detection approach which was proposed by Paul Viola and Michael Jones in their paper, **"Rapid Object Detection using a Boosted Cascade of Simple Features"** in 2001.

This project recognizes the facial expressions of user and play songs according to emotion. Facial expressions are best way of expressing mood of a person. The facial expressions are captured using a webcam and face detection is done by using Haar cascade classifier.

The captured image is input to CNN which learn features and these features are analyzed to determine the current emotion of user then the music will be played according to the emotion. In this project, five emotions are considered for classification which includes happy, sad, anger, surprise, neutral. This project consists of 4 modules-face detection, feature extraction, emotion detection, songs classification. Face detection is done by Haar cascade classifier, feature extraction and emotion detection are done by CNN. Finally, the songs are played according to the emotion recognized.

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, such as images, which is inspired by the organization of animal visual cortex 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 of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output, such as classification.

A convolution layer plays a key role in CNN, which is composed of a stack of

mathematical operations, such as convolution, a specialized type of 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.

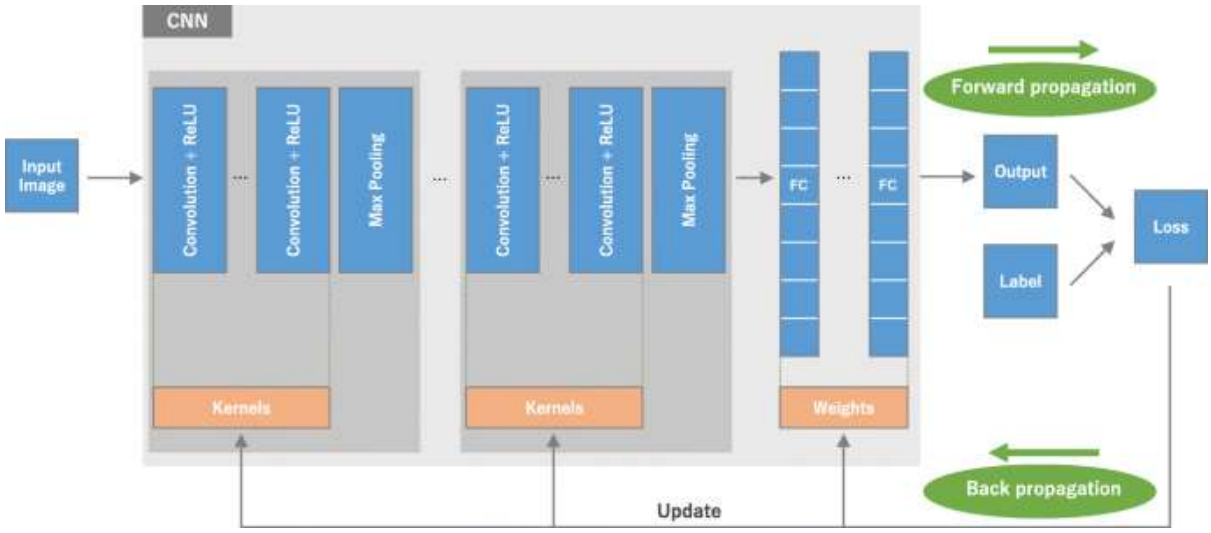

Fig. 1: Basic CNN

Applications of Computer Vision:

1. Autonomous Vehicles.
2. Facial Recognition.
3. Image Search and Object Recognition.

Advantages of Computer Vision:

1. Faster and simpler process
2. Better products and services
3. Cost-reduction

Disadvantages of Computer Vision:

1.1 MACHINE LEARNING

Machine Learning is the most popular technique of predicting or classifying information to help people in making necessary decisions. Machine Learning algorithms are trained over instances or examples through which they learn from past experiences and analyze the historical data. Simply building models is not enough. You must also optimize and tune the model appropriately so that it provides you with accurate results. Optimization techniques involve tuning the hyperparameters to reach an optimum result. As it trains over the examples, again and again, it is able to identify patterns in order to make decisions more accurately. Whenever any new input is introduced to the ML model, it applies its learned patterns over the new data to make future predictions. Based on the final accuracy, one can optimize their models using various standardized approaches. In this way, Machine Learning model learns to adapt to new examples and produce better results.

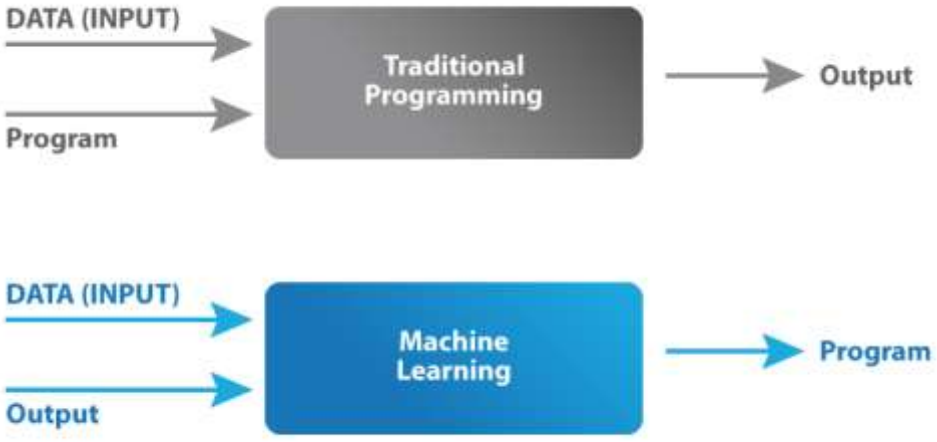

Fig. 2: Traditional Programming and Machine Learning

TYPES OF LEARNINGS:

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

1. Supervised learning
2. Unsupervised Learning
3. Reinforcement Learning

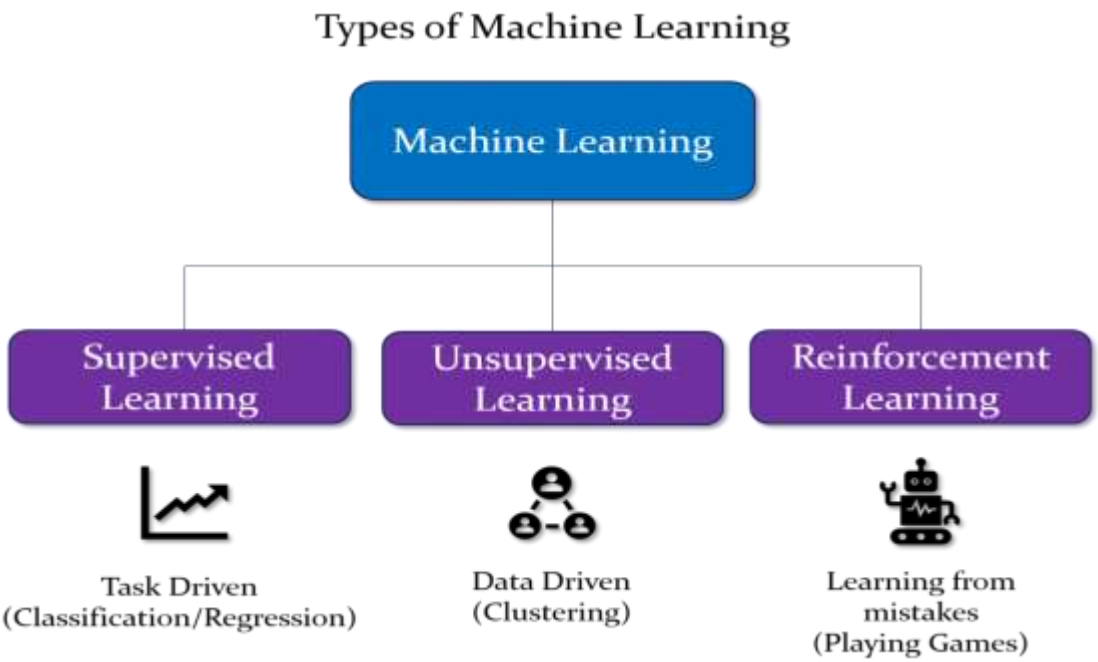

Fig. 3: Types of Machine Learning

SUPERVISED LEARNING:

Supervised learning is the most popular paradigm for machine learning. It is the easiest to understand and the simplest to implement. It is the task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. Supervised Learning is very similar to teaching a child with the given data and that data is in the form of examples with labels, we can feed a learning algorithm with these example-label pairs one by one, allowing the algorithm to predict the right answer or not. Over time, the algorithm will learn to approximate the exact nature of the relationship between examples and their labels. When fully trained, the supervised learning algorithm will be able to observe a new, never-before-seen example and predict a good label for it.

Most of the practical machine learning uses supervised learning. Supervised learning is where you have input variable (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(x)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for the data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. Supervised learning is often described as task oriented. It is highly focused on a singular task, feeding more and more examples to the algorithm until it can accurately perform on that task. This is the learning type that you will most likely encounter, as it is exhibited in many of the common applications like Advertisement Popularity, Spam Classification, face recognition.

Two types of Supervised Learning are:

1. Regression:

Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Regression can be used to estimate/ predict continuous values (Real valued output). For example, given a picture of a person then we have to predict the age on the basis of the given picture.

2. Classification:

Classification means to group the output into a class. If the data is discrete or categorical then it is a classification problem. For example, given data about the sizes of houses in the real estate market, making our output about whether the house "sells for more or less than the asking price" i.e., Classifying houses into two discrete categories.

UNSUPERVISED LEARNING

Unsupervised Learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabeled data and looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. In contrast to supervised learning that usually makes use of human labelled data, unsupervised learning, also known as self-organization, allows for modelling of probability densities over inputs.

Unsupervised machine learning algorithms infer patterns from a dataset without reference to known, or labelled outcomes. It is the training of machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, machine is restricted to find the hidden structure in unlabeled data by our-self. For example, if we provide some pictures of dogs and cats to the machine to categorized, then initially the machine has no idea about the features of dogs and cats so it categorizes them according to their similarities, patterns and differences. The Unsupervised Learning algorithms allows you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning methods.

Unsupervised learning problems are classified into two categories of algorithms:

- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

REINFORCEMENT LEARNING

Reinforcement Learning (RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. Machine mainly learns from past experiences and tries to perform best possible solution to a certain problem. It is the training of machine learning models to make a sequence of decisions. Though both supervised and reinforcement learning use mapping between input and output, unlike supervised learning where the feedback provided to the agent is correct set of actions for performing a task, reinforcement learning uses rewards and punishments as signals for positive and negative behaviour. Reinforcement learning is currently the most effective way to hint machine's creativity