FACIAL EMOTIONS BASED SONG RECOMMENDATION

Dr.I.Selvamani[1]., Mukkala Priyanka[2]., Sirigireddy Ruchitha[3]., Srirangaraju Sai Varsha[4]., Kancheti Harshitha[5]

1 Professor, Department Of ECE., Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India (✉i.selvamani@gmail.com)

2,3,4,5 B.Tech ECE, (19RG1A04G0, 19RG1A04G6, 19RG1A04G8, 19RG1A04H6),

Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

Abstract

To automatically play music based on face expression, we offer a novel method. Most current methods either need human intervention in the form of playing music or the use of wearable computers or audio feature classification. Instead, we suggest automating the selection and playing processes. For this purpose, we used a Convolutional Neural Network. Pygame and Tkinter are used to provide playlist suggestions. In general, our suggested approach improves the accuracy of the specified system while decreasing the amount of time and money needed to compute the results. The FER2013 dataset is used for testing the system. An inside camera records the user's facial expressions. In order to identify feelings such as happiness, anger, sadness, surprise, and neutrality, feature extraction is applied to the input facial photos. A user's current mood is used to mechanically curate a musical mood-board. In terms of computing time, it outperforms the approach used in the current literature.

Keywords

Keywords: Convolutional Neural Networks, Pygame, Tkinter, Face Recognition, Feature Extraction, Emotion Detection, Player, Camera.

1. Introduction

Recent research have shown that music has a profound effect on people's emotional states and mental processes. Music was shown to have a significant effect in linking arousal and mood in an analysis of the reasons why individuals listen to music. Participants evaluated music highly for its potential to improve their mood and increase their self-awareness. There is strong evidence linking musical tastes to character characteristics and emotional states [1].

Emotional and behavioral circuits are responsible for controlling musical aspects such as tempo, timbre, rhythm, and pitch [2]. It's possible that social interaction is an integral part of your daily routine. Whether via body language, voice, facial expression, or emotional state, it displays flawless details and a great deal of data among people [3]. Smart card applications, surveillance, image database investigation, criminal, video indexing, civilian applications, security, and adaptive human-computer interface with multimedia environments are just some of the many areas where emotion detection has found widespread use recently.

Automated emotion detection in multimedia attributes like music or movies is on the rise thanks to developments in digital signal processing and other efficient feature extraction algorithms, and it has the potential to contribute significantly to fields as diverse as human-computer interaction systems and music entertainment. We propose an emotion-recognition recommender system [13–24] that uses facial expressions to identify the user's emotional state and provide a selection of music that would effectively convey that state. The suggested system is able to pick up on a person's feelings, and if he or she is feeling down, a playlist with the most appropriately upbeat songs would be presented. And if the feeling is one of joy, a unique playlist will be played that includes a variety of upbeat songs [4].

We utilized the Kaggle Facial Expression Recognition [5] dataset to identify emotions. The player's data set is based on a compilation of Bollywood and Hindi songs. Convolutional Neural Network is used to implement facial emotion recognition, and it has an accuracy of 95.14 percent [2].

2. Literature Review

The purpose of this analysis is to better understand the methodologies and the challenges they provide so that we may find solutions. Literature reviews, also known as literature surveys, are sections of academic papers that summarize the state of knowledge on a certain issue and highlight significant new results, as well as theoretical and methodological advancements. Learners, scientists, engineers, and others from all around the globe have taken an interest in the innate abilities of people because of the many ways in which they might contribute to the inputs of any given system.

Facial expressions reveal the individual's present mental state. When communicating with others, we almost always rely on nonverbal cues such as hand gestures, facial expressions, and tone of voice to convey our emotions. According to Preema et al. [6], creating and maintaining a sizable playlist is a tedious and time-consuming process. According to the research study, "the music player itself chooses a song based on the user's current mood." The

software analyzes the audio files and sorts them into genres and mood categories. The Viola-Jonas algorithm, which can identify faces and extract emotions, is used by the program. To classify the collected characteristics into the five most common human emotions—anger, pleasure, surprise, sadness, and disgust—a Support Vector Machine (SVM) was utilized.

In their study [3], authors Yusuf Yaslan et al. propose a wearable computing device-integrated galvanic skin response (GSR) and photoplethysmography (PPG) physiological sensor-based emotion-based music selection system. Feelings are an integral element of being human. They're important in all aspects of living. In this work, we consider the emotion recognition issue as the prediction of arousal and valence from many physiological channels. According to research by Ayush Guidel, et al., [7] facial expressions provide a clear window into a person's mental and emotional condition. The development of this approach took the eight primary emotions into account: happiness, sadness, anger, excitement, surprise, disgust, fear, and apathy. In this research, we used a convolutional neural network to recognise faces. Music is sometimes described as a "language of emotions" among cultures all over the world.

The intelligent music player that takes emotion detection into account was suggested in a work by Ramya Ramanathan et al [1]. Humans are wired to feel emotions. They're crucial in every aspect of existence. Feelings are designed to be communicated and understood by others. The user's own music library is originally organized by the mood each record is meant to evoke. This is frequently determined by analyzing the words to the music. In the publication, the authors argue that

The focus of this paper is on the uniqueness of the existing methods for emotion recognition, their application to the development of emotion-based music players, and the optimal implementation of the suggested system for emotion detection. It also provides a high-level overview of how our systems function, how we generate playlists, and how we classify emotions.CH Radhika et al. [8] suggested that it is laborious and time consuming to manually segregate a playlist and annotate music according to a user's current emotional state. Various methods for doing so have been suggested. However, current algorithms are sluggish, offer significantly less precision, and raise the total cost of the system by making use of extra hardware (such as EEG structures and sensors). This work introduces an algorithm that can automatically generate an audio playlist based on a person's facial expressions, saving the user the time and effort of doing it manually. The study provides an algorithm that aims to cut down on the total computational time and cost of the planned system. It also aspires to improve the precision of the system's blueprint. The system's face expression recognition component is tested and verified using a dataset that includes data from both human and nonhuman subjects.

3. Problem Definition

Create an app that uses a user's facial expressions captured by a webcam to provide personalized music recommendations in real time.

4. Proposed System Overview

- We may gain from the suggested system by introducing user-music player interaction. The system's goal is to take high-quality photographs of people's faces. The Convolutional Neural Network is then used to make an emotional prediction based on the input photos. After extracting feelings from the photo, a musical selection is made. Our suggested system's primary function is to instantly alter the user's mood by playing music that either makes them happy, sad, natural, or shocked. If the suggested system identifies that a subject is experiencing a negative feeling, for example, that individual will be sent to a playlist containing the most effective types of music for lifting their spirits. There are four parts to face expression recognition-based music recommendations.
- • Real-Time Capture: This section requires accurate facial capture while the user is actively using the device.
- In the case of "Face Recognition," the user's face will be used as input. The convolutional neural network is trained to analyze the user's picture for certain characteristics.
- Caption generation is determined by the user's emotional state, which is determined by extracting characteristics from the user's picture in the "emotion detection" portion.
- The user's feelings are mapped into the song's mood, and then the recommendation module suggests a song to the user.

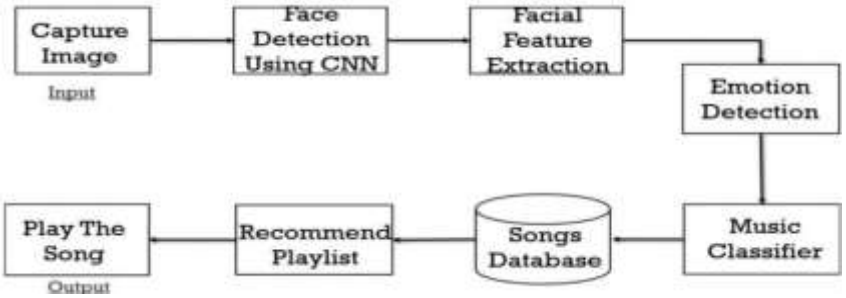


Figure 1. Block diagram of the proposed system.

5. Methodology

5.1 Database Description

Using the Kaggle dataset, we constructed a Convolutional Neural Network model. The dataset, FER2013, is divided into a training set and a test set. There are a total of 24176 images in the training dataset, whereas there are 6043 in the testing dataset. The collection contains grayscale face pictures with a resolution of 48x48 pixels. Five different emotions (happy, sad, furious, surprised, and neutral) are assigned to each picture in FER-2013. Automatic face registration ensures that the faces are roughly centered and occupy the same amount of space throughout all images. There are both posed and candid photos of people's faces in FER-2013, all of which are grayscale and 48x48 pixels in size.

By compiling the outcomes of a Google image search for each emotion and their synonyms, the FER-2013 dataset was generated. If a FER system is trained with an unbalanced dataset, it may do well with more common emotions like happiness and sadness, but struggle with less common ones like fear and contempt. The common method for dealing with this issue is the weighted-SoftMax loss method, which assigns a different weight to the loss term for each emotion class based on its representation in the training set. However, the SoftMax loss function, upon which this weighted-loss strategy is based, is said to readily drive features from different classes to remain separate without taking intra-class compactness into account. Using an auxiliary loss to guide the neural network is one way to tackle the issue of SoftMax loss. We have utilized a loss function called categorical crossentropy to deal with missing and Outlier variables. A different loss function is used to quantify the mistake at each iteration. Therefore, we have utilized a loss function called categorical crossentropy to deal with missing and Outlier data.

.

Figure 2. Samples from FER2013 dataset.

5.2 Emotion Detection Module

5.2.1 Face Detection

Computer vision technology has several potential uses, one of which is face identification. Creating and training algorithms for accurate face or object identification in an object detection or related system
images. This identification may occur in real time, directly from an image or video frame. Classifiers, algorithms that determine whether an object in a picture is a face (1) or not a face (0), are used for face detection. To improve their accuracy, classifiers are taught to use a large dataset of photos to learn to recognize faces. Local Binary Pattern (LBP) and Haar Cas- cas are two types of classifiers that are implemented in OpenCV. For reliable face identification, researchers have turned to the Haar classifier, a model that can be trained using a variety of pre-defined face data. The primary goal of face detection is to isolate the face in the picture from other distracting elements. This method uses machine learning to train the cascade function on a set of input files. The Haar Wavelet method is used to analyze the image's pixels into squares according to their function [9]. This makes use of machine learning methods to encourage precision using "training data."

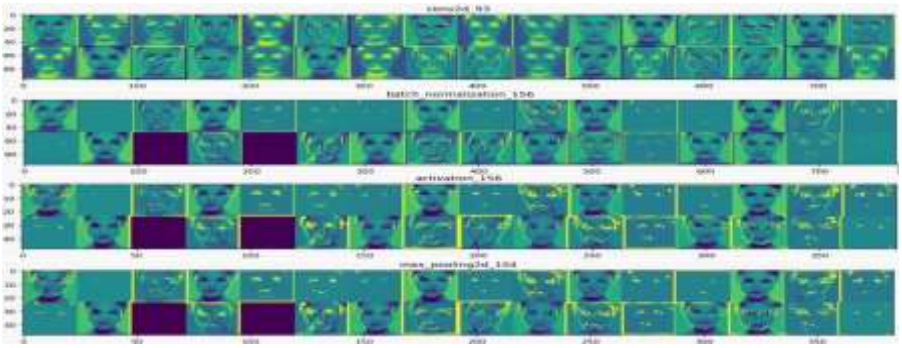
".

Figure 3. Face detection.

5.2.2 Feature Extraction

We use the pre-trained network, which is a sequential model, as if it were a generic feature extractor when we execute feature extraction. By letting the input picture flow through to the next layer, and then halting there, we may use the features produced by that layer. High-level features are extracted from the captured picture by the first layers of a convolutional network, therefore keep the number of filters low. For more dense layers, we use filters with twice or treble the resolution of their predecessors. The deeper the layer, the more features the filter can pick up, but the more processing is required.

Figure 4. Visualization of The Feature Map.



Here, we made use of the Convolution neural network's [10] ability to train strong, discriminatory features. The model's outputs will be feature maps, which serve as a representational intermediary for layers beyond the first. It is necessary to load the input picture in order to examine the Feature map and determine which characteristics were most important in classifying the image. Applying Filters or Feature detectors on the input picture or the feature map output of the preceding layers yields the resulting feature map. The internal representations for distinct inputs for each Convolutional layer in the model may be better understood with the help of a feature map visualization..

5.2.3 Emotion Detection

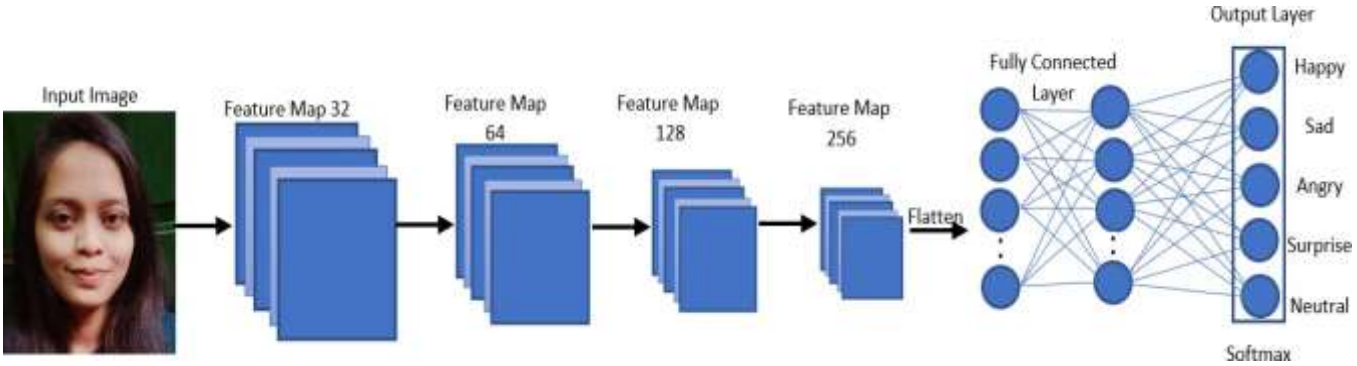


Figure 5. Convolution neural network Architecture.

The Relu activation function [11] is applied to the input picture using filters or feature detectors in a convolutional neural network architecture, yielding feature maps or activation maps. Edges, lines, curves, and other features in a picture may be detected and isolated with the use of feature detectors and filters. The feature maps are then pooled to make them translation-independent. The assumption behind pooling is that the pooled outputs remain stable in the face of small changes to the input. Choose minimum, median, or maximum pooling at your leisure. However, max-pooling outperforms min-pooling and even average-pooling. Inputs are flattened and sent to a deep neural network, the outputs of which are the object's class.

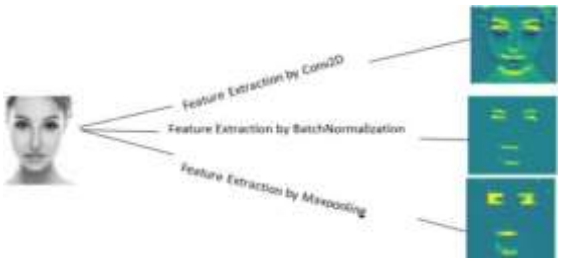


Figure 6. Feature Extraction by each layer in Convolutional Neural Network.

The class of the image will be binary, or it will be a multi-class classification for identifying digits or separating various ap- parel items. Neural networks are as a black box, and learned features in a Neural Network are not interpretable. So basically, we give an input image then the CNN model returns the results [10].

Emotion detection is performed by loading the model which is trained by weights using CNN. When we take the real-time image by a user then that image was sent to the pre-trained CNN model, then predict the emotion and adds the label to the image.

Figure 7. Results of Emotion Detection.

5.3 Music Recommendation Module

5.3.1 Songs Database

We assembled a repository of Bollywood-style Hindi music. Depending on the feeling, there are between a hundred and a few hundred songs included. We can all agree that music plays a significant role in lifting our spirits. If a person reports feeling down, the system will provide a set of musical selections designed to lift their spirits..

5.3.2 Music Playlist Recommendation

The user's mood may be gauged in real time with the help of the emotion module. Labels like "Happy," "Sad," "Angry," "Surprise," and "Neutral" will be generated. Our custom-built music database's folders are linked to their respective labels through the os.listdir() function in Python. The songs are listed in Table 1. To get a list of all files located in the given directories, use the os.listdir() function.
os.chdir("C:/Users/deepali/Downloads/Happy") if label=='Happy' self.mood.set("You are looking happy, I am playing song for You");

os.listdir() songtracks = # Obtaining Music

Table 1. Database of songs.

Emotion	Songs
Happy	Track 1"Dil Dhadakne Do"
	Track 2"Aaj Mai Upar"
	Track 3 "Ilahi"
Sad	Track 1 "Apna Time Aayega"
	Track 2 "Ruk Jana Nahi"
	Track 3 "All is Well"

Angry	Track I "Dushman Na Kare Dost Ne"
	Track 2 "Thukra Ke Mera Pyaar"
	Track 3 "Khalbali'
Surprise	Track I "Zindagi Kaisi Hai Paheli"
	Track 2 "Aao Milon Chalen"
	Track 3 "Jaane Kyun"
Neutral	Track I "Buddhu Sa Mann"
	Track 2 "Matargashti"
	Track 3 "Dildara"

```
# Inserting Songs into Playlist
        for track in songtracks:
            self.playlist.insert (END, track)
```

This will cause the music player's user interface to propose a playlist to the user based on their current emotional state. Since Pygame is capable of playing many different multi-media formats, we've utilized it to play the audio. Playsong, pausesong, resumesong, and stopsong are all useful functions of this library when interacting with the music player. The names of all songs, the state of currently playing songs, and the primary GUI window are all kept in separate variables called playlist, songstatus, and root, respectively. We've utilized Tkinter for the UI development.Here are the results:

Figure 8. GUI of the front page.

Figure 9. Detection of emotion.

Figure 10. Recommendation of music playlist.

6. Result & Analysis

We analyzed many papers that used convolutional neural networks, extreme learning machines, and support vector machines [12]. Table 2 compares several similar algorithms to one another. Each research has its own unique set of algorithms and accuracy metrics. Emotion detection accuracy is enhanced when a Convolutional Neural Network is used..

Table 2. Validation and Testing accuracy for the three algorithms on the Fer2013 Dataset.

Algorithm	SVM	ELM	CNN
Validation Accuracy	0.66	0.62	0.95
Testing Accuracy	0.66	0.63	0.71

Table 3 displays the trained CNN's hyperparameters. The weight is updated at the conclusion of each batch, and this is controlled by the learning rate. Train the network across many epochs using the whole training dataset. The batch size is the number of iterations that occur before the network's weights are modified. Nonlinear prediction bounds may be learned by the model with the help of activation functions. Adam might replace stochastic gradient descent as an alternative optimization technique in deep learning model training. In single-label, multi-class classification tasks, the loss function categorical- crossentropy is used to quantify deep learning model err- rors..

Table 3. Hyperparameter for trained CNN network.

Hyperparameters	Values
Batch size	128
No. of classes	5
Optimizer	Adam
Learning rate	0.001
Epoch	48
No. of Layers	28
Activation function	Relu, SoftMax
Loss function	Categorical-crossentropy

7. Conclusion

Research shows that there are several ways to create a music recommendation system. The techniques offered by researchers and creators in the past were investigated. The results helped us solidify the system's goals. With the growing popularity of AI-driven apps comes the opportunity for our project to make use of cutting-edge technology. In this system, we explain how music affects the user's emotions and how to choose the best songs to lift the user's spirits. The user's moods can be read by the implemented system. The method was able to identify five distinct emotions: joy, sadness, anger, apathy, and surprise. The suggested technology identified the user's mood and then presented them with a playlist of songs that matched their emotional state. It takes a lot of memory and processing power to work with a large dataset. This will increase the difficulty and allure of development. The goal is to build this app using standardized tools and at the lowest feasible cost. With our face expression recognition-based music recommendation system, consumers will spend less time curating and organizing their music libraries..

8. Future Scope

While it is fully functional at the moment, there is room for development in this system. The program may be improved in many ways to provide the user with a more satisfying experience. Some of them use an other approach, one that takes into account feelings like revulsion and fear that aren't included in our standard model. As part of this feeling, we've included in the ability to play music automatically. The system's long-term potential will allow for the development of a mechanism that might be useful in music therapy, therefore facilitating the treatment of patients with conditions such as stress, anxiety, acute depression, and trauma. As such, there is room for improvement in the existing system, which struggles under severe low-light circumstances and low-resolution camera setups.

Acknowledgement

A B.E. project can't be built without the help and direction of many individuals. Therefore, we think it is of the utmost importance to express our gratitude to everyone who assisted us in this endeavor.

We appreciate Dr. P. D. Deshmukh, our principal, for encouraging us and providing us with the tools we needed to complete this assignment. We'd also like to thank Dr. Sanjay Sharma, head of the Computer Engineering department, for his helpful cooperation.

Ms. Megha V. Gupta, our Project Guide, was an invaluable resource during the whole process of developing this project, and we are grateful to her for her insightful comments, continual guidance, and positive reinforcement.

Finally, we would want to express our appreciation to the friends, faculty, and staff who helped us improve our B.E. Project via their words of support and constructive criticism. For their unwavering love and encouragement over the years, we also appreciate our parents.

References

[1] An intelligent music player using emotion detection, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2017, Ramya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu. https://doi.org/10.1109/CSITSS.2017.8447743

[2] Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, (IEEE), 2017; Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, Smart music player incorporating face expression identification and music mood suggestion. https://doi.org/10.1109/WiSPNET.2017.8299738

[3] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak. "Emotion-based music recommendation system using wearable physiological

[4] May 2018 issue of IEEE Transactions on Consumer Electronics, page(s):14-8, sensors.https://doi.org/10.1109/TCE.2018.2844736

[5] Kholood Albalawi, Liyakathunisa Syed, Ahlam Alrihail, Alaa Alsaedi, Department of Computer Science, Taibah University (DeSE), 2019, Emotion-based music recommendation system for users based on face feature analysis. https://doi.org/10.1109/DeSE.2019.00188

[6] [5] Challenges in representation learning, with an emphasis on face expression recognition, Research Prediction Competition Recognize emotions based on a photograph (KAGGLE).

[7] Sahana M, Savitri H, Preema J.S., Rajashree Analysis of a music player that responds to facial expressions, Volume 6, Issue 15, 2018, International Journal of Engineering Research and Technology (IJERT), ISSN: 2278-0181.

[8] Guidel, Birat Sapkota, and Krishna Sapkota of the AYUSH, Facial recognition music recommendations for February 17, 2020.

[9] The following article is in the April 2020 issue of the International Journal of Emerging Technologies and Innovative Research (JETIR): [8] CH. sadhvika, Gutta.Abigna, P. Srinivas reddy, Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad.

[10] Becoming Human AI (2019), Vincent Tabora, Face

Detection Using OpenCV and Haar Cascade Classifiers.

[11] As cited in [10] Zhuwei Qin, Fuxun Yu, Chenchen Liu, and Xiang Chen. What the world looks like to a convolutional neural network: an overview of approaches to visualizing CNNs. Computer Science and Its Mathematical Foundations, May 2018.

[12] [11] Using a Real-Time Camera for Emotion Detection in an Emotion-Based Music Player, Ahmed Hamdy AlDeeb, ResearchGate, June 2019.

[13] [12] Authors: Frans Norden and Filip von Reis Marlevi TRITA-EECS-EX-2019:143, A Comparative Study of Machine Learning Algorithms for Detecting Two-State Changes in Facial Expressions.

[14] [13] To interpret and localize Thorax illnesses using DCNN in Chest X-Ray, P. Singhal, P. Singh, and A. Vidyarthi (2020). 1(1), 1-7, Journal of Electrical and Electronics Engineering and Computer Science

[15] [14] BlueBrain, a review of artificial intelligence technology by M. Vinny and P. Singh, is due out in 2020. 1-3 in the first issue of the Journal of Information, Electrical, and Electronics Engineering.

[16] Detecting objects is the topic of [15] A. Singh and P. Singh (2021). Management and Service Science Journal, Volume 1 Number 2 Issue 3 (pp. 1-20).

[17] Referring to: [16] Image Classification: A Survey by A. Singh and P. Singh (2020). Research in Electrical and Electronics Engineering, 1(2), 1-9, Journal of Informatics.

[18] License Plate Recognition, by A. Singh and P. Singh (2021). Management and Service Science Journal, Volume 1 Number 2 (January 2011), Pages 1-14.