Intelligent Music Recommendation System based on Face Emotion Recognition

2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 979-8-3503-0611-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCIS60361.2023.10425186

Nandini Gupta
Department of CSE
Graphic Era Deemed to be University,
Dehradun, India
nandinigupta268@gmail.com

Shruti Agarwal
Department of CSE
Graphic Era Deemed to be University,
Dehradun, India
shrutiagarwal0444@gmail.com

Kireet Joshi
Department of CSE
Graphic Era Deemed to be University,
Dehradun, India
joshikireet@gmail.com

Vishan Kumar Gupta
Department of CSE
Amity University, Punjab
Mohali, India
vishangupta@gmail.com

Surendra Kumar Shukla
Department of Computer Engineering
SVKMS NMIMS MPSTME,
Shirpur Campus, India
surendrakumar.shukla@nmims.edu

Gurpreet Singh
Department of CSE
Sir Padampat Singhania University,
Udaipur, India
gurpreet.singh@spsu.ac.in

***Abstract*— Music has been considered a form of expression that can treat conditions such as stress, anxiety, and mood swings. The latest research in music psychology shows that music evokes palpable emotions in listeners. People's personality and mood determines what kind of music they prefer to listen to. The areas of the brain that process thoughts and emotions also control tempo, timbre, and music. Although a person's response to music depends on many external factors, we can classify music as kindness, patience, relaxation, and passion, that's why music is a high-quality connector. Industries are also part of us for a long time, backgrounds, languages, hobbies, political leanings and earning degrees, songs, and other streaming apps. Those apps may be used every time and everywhere and they can be used for daily activities, travel, sports, activities, etc. It is an excessive call for as it can be combined with the rapid development of cellular networks and the digital multimedia era, the virtual track has ended up being the main content of many younger human beings. Emotion-based tunes in our device is designed to recognize emotions in actual time and suggest songs primarily based on perceived emotions. This will become an addition to the conventional tune participant apps we had on our telephones before. One of the essential blessings of being concerned in emotional intelligence is purchaser satisfaction. The main motive of this device is to research the user's image, predict the expression of the person and advise songs appropriate to the detected mood.**

Keywords- motion recognize, facial extraction, music recommendation, convolutional neural network, web camera, machine learning.

I. INTRODUCTION

As we know, Spotify, Gaana etc., are already many music recommendation systems, all based on research questions rather than the user's current preferences. To improve current needs, we can enable psychology and technology to analyse emotions and find beauty for users. To detect facial expressions, we can divide people's emotions into five emotional categories such as fear of person, disgust, anger of person, happy faces, sadness of person's face, surprise faces and neutral [1]. For this work, we have divided the songs into five basic moods: happy, sad, neutral, angry, surprised. Since it is important to recommend music to users based on their current mood, we will create a list to recommend after analysing their feelings and preferences. Recent research on emotions based advice often focuses on two issues, music, and sound. To break the barrier of the language our motive is to focus on how we will extract the sound and how we will analyse the hindi songs and then we will share these features with the emotions of five people and then create playlists accordingly.

II. LITERATURE SURVEY

P. Rajashree, M. Sahana, H. Savitri proposed in his paper a song player that with its own select's music in line with the current mood of the user [1]. Their software makes use of Viola- Jonas's algorithm for face detection and facial features extraction. It additionally makes use of help Vector device set of rules for category of extracted functions into five basic feelings like anger, disappointment, disgust, pleasure, and surprise. D. Reney and N. Tripathi [2] of their paper an efficient approach to standard and emotion detection have detected face from the enter photo using Viola Jonas face detection algorithm after which detected the emotion of the face the use of K-Nearest buddies' classifier. Zeng et al [3] proposed a paper wherein he classes facial feature into predominant classes, which protected appearance based totally function extraction and geometric primarily based characteristic extraction, which protected extraction of some critical factors of the face which include mouth, eyes, eyebrows. In the paper of Shantha Shalini. K. et al [4], facial emotion-based tune advice device using pc vision and device gaining knowledge of techniques in this proposed system, laptop vision and machine mastering strategies are used for connecting facial emotion for tune recommendation. The capabilities from the enters snap shots are extracted using a factor detection set of rules. The type set of rules OpenCV is used for training the input images for facial

ISBN: 979-8-3503-0611-8/23/$31.00 ©2023 IEEE

emotion detections. Yadhukrishna, Rajalakshmi, and Prithviraj [5] proposed "facial features recognition using CNN and LBP," 2020 IEEE, this paper hopes to carry an honest contrast of the two most normally used face expression reputation [FER] techniques and shed mild on their accuracy. These techniques are used local binary styles [LBP] and convolutional neural networks [CNN] [6].

III. OVERVIEW

Humans often express their feelings by their expressions, hand gestures, and by raising the voice of tone but mostly humans express their feelings by their face. Emotion-based music player reduces the time complexity of the user. Usually, people have many songs on their playlist [7]. Playing songs randomly does not satisfy the mood of the user. This system helps users to play songs automatically according to their mood. The suggested system captures the emotion of the person and then it'll create a playlist of all the songs recommended to the user according to their facial emotions and preferences [8]. The following modules need to be implemented:

1) *Facial Emotion Recognition Module:* This will recognize the emotions based on the extracted facial feature and classify into five emotions.
- Developing the facial emotion recognition model.
- Evaluate the model by capturing the image.

2) *Music Recommendation Module:* This will recommend songs like the previous songs listened to by user. This will take into consideration the song features and the emotion displayed by the user.
- Developing the music classifier model.
- Classifying the user's playlist according to emotions

IV. PROPOSED METHODOLOGY

A. About Dataset

For the emotion category we will use FER 2013 as our dataset consisting of photos portraying various facial expressions. The pics are labelled into five categories: happy, sad, anger, neutral, surprised. The dataset called FER 2013 comprises monochrome photographs of faces that measure 48*48 pixels [9]. The faces have been mechanically registered in order that the face is more or much less targeted and occupies approximately the identical amount of area in each photograph. The project involves assigning each face to a particular emotional category, solely based on the expression displayed by their facial features. These categories include anger (0), happy (1), sad (2), surprise (4), neutral (5). The training set includes 24,176 photographs and the validation set includes 6,043 photographs [10]. The total number of images for each emotion are:

- 3995 images of angry faces
- 7215 images of happy faces
- 4965 images of neutral faces
- 4830 images of sad faces
- 3171 images of surprise faces

Fig. 1. Samples of different facial emotions (angry, happy, neutral, sad, and surprise)

B. Preparing the Data

The first step involves data augmentation, a technique that promotes expansion in the training datasets image count. The flow from directory class in the Image Data Generator reads images from folders that contain images. The images will be loaded using it [11].

C. Developing the CNN Model

We use neural networks to train the model to recognize facial expressions. This step is the most important part of the whole process when we create a CNN where we will pass our features to train the model and finally test it with benchmarks. CNN is a neural network often used in image recognition and processing because of its ability to recognize patterns in images [12].

The layers that make up the CNN are of three types: convolutional layers, pooling layers, and fully connected layers (FC). When these layers are combined, the CNN architecture is created. In addition to these three methods, there are two other important methods namely release and activation methods. Edges, corners etc. Input, convolution, rectified linear units (RELU), pooling, and fully connected layers make up the Convolutional Neural Network and M filters act as straightforward feature extractors [13].

The first layer, known as INPUT, contains untouched pixel data, as implied by its name. The information contained within an image can be referred to as its raw pixel's values. An instance of this can be seen in an INPUT [64*64*3], which indicates an image with dimensions of 64 pixels in width, 64 pixels height and three layers of colour representing red, green and blue.

The convolution layer serves as the fundamental component of CNNs since the majority of calculations take place within it .

For instance, by utilizing the [6*4*64*3] filters provided in the table, we can produce a volume of [64*64*6].

The activation function of the previous layer is used in the output of the layered layer, which is commonly referred to as RELU. To put it differently, the RELU network will be modified to include nonlinearity [14].

POOL refers to a layer which is a fundamental component of CNNs, known as layers. This method operates on each individual input segment and modifies its spatial size with the primary goal of reducing sampling rate [15]. FC is known as the full link layer or a layer higher than the release layer.

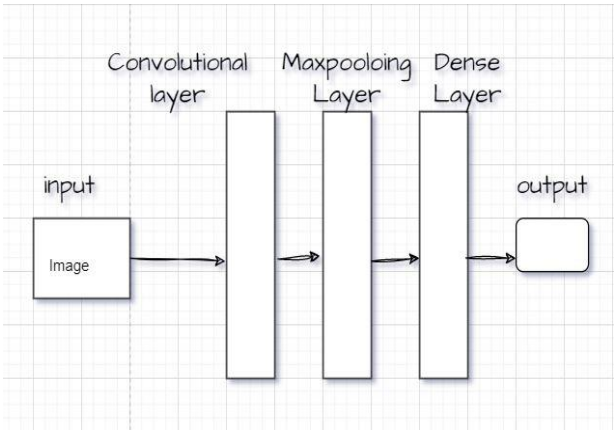

Fig. 2. CNN architecture

The input shape of the images is (48,48,1), where 48, 48 are the height and width of the images. Number 1 signifies the colour channel, but for grayscale images, it is only one, whereas for RGB images, it is three. To create the model, we utilized the adaptive learning rate technique known as Adam, evaluated losses during training using binary cross entropy, and assessed the model's performance with accuracy.

Training the Model: The model is being given certain parameters. The model utilizes a train set with 50 epochs, a validation set to compute both Val loss and Val accuracy. Along with class weights and a callback list [16].

D. Evaluating the Model

After running our code to train the model on training dataset the training accuracy comes out to be 91.86% as depicted Figure 3 and Figure 4.

```
#saving values for confusion matrix and analysis
np.save('truey', truey)
np.save('predy', predy)
print("Predicted and true label values saved")
print("Accuracy of the model :"+str(acc)+"%")
```

```
Loaded model from disk
225/225 [==============================] - 20s 79ms/step
Predicted and true label values saved
Accuracy of the model :91.90582334912231%
```

Fig. 3. Accuracy of the model

E. Designing of Face Capture and Preprocessing System

The process of recognizing faces is commonly explained as a procedure that initially encompasses four stages: namely, detecting faces, aligning them, extracting notable features, and ultimately recognizing the face [17].

- *Face Detection Process:* Find one or more multiple faces in the picture and indicate their location using a bounding box.
- *Face Alignment Process:* Make the face conform to the database by ensuring its geometry and photometrics are in line with it.
- *Feature Extraction Process:* Make facial characteristics that are applicable for the purpose of identification.
- *Face Recognition Process:* Compare the face to one or more well-known faces in the database that has already been created [18].

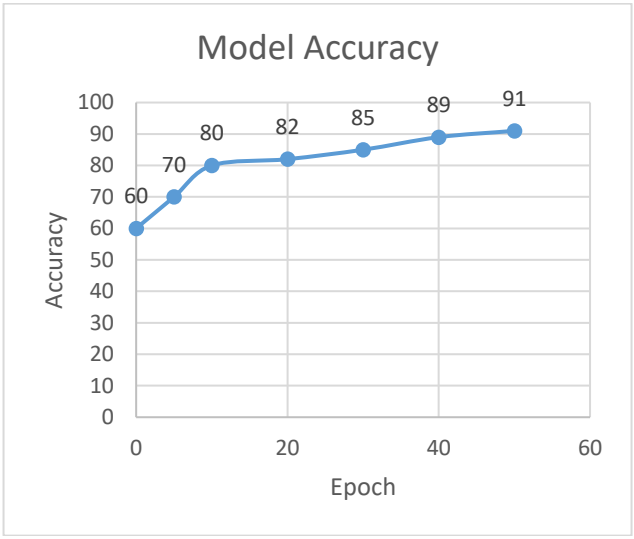

Fig. 4. Training and accuracy outcome

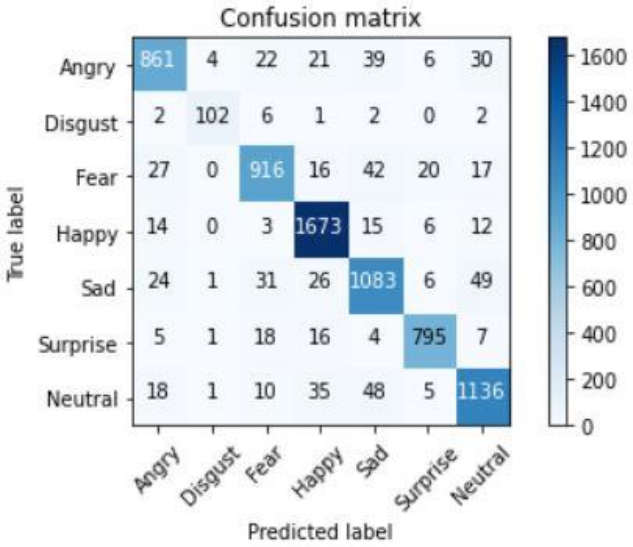

Fig. 5. Confusion matrix

F. Designing of Music Classifier System

When the convolution process is learned from many sound spectra, it is now possible to obtain specific maps with a high level of abstraction. The mapping function can be expanded over time to change the order of the features and feed the order of the joint to the music processing model BiRNN [19]. Then, using the weight of the focus, the network is trained to collect the content and weight of the results from the Bi-RNN and combine the output of the BiRNN several times and turn it into a musical synthesis process.

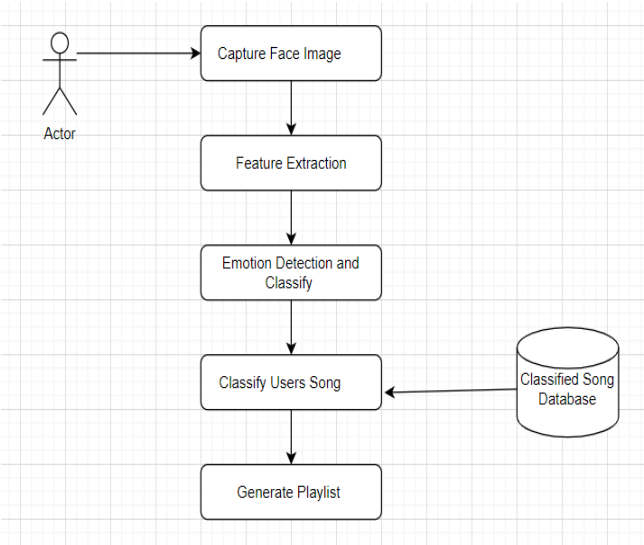

Fig. 6. Flow of proposed work.

G. Music Recommendation

The input image is taken from a webcam for live view capture. Here are the four key points of, because it's hard to pinpoint all the different behaviors, and by using fewer options, helps with writing time and results are more difficult. Compares the current value against the threshold in code [20]. These results will be submitted for study on the website. The song will be played by perceived emotions. Thoughts are given to each song. When requests are submitted, the corresponding songs and characters are counted and assigned to each song. However, we can use many models to represent because of its accuracy. We are using fisher face which has PCA and LDA algorithms, so it gives better accuracy than other algorithms. For sound mechanics, we use mechanics, a python library for sound gain and frequency and time. The result is compared with the current value as a starting point. Besides the thought-based system of queue mode and random mode, there are other options to choose from. In line, we can make playlists like other music software, the last one is random mode which choose songs instead of ordering food [20]. When the song is played on, it represents that mood in four different modes in the form of emotions, depending on the mood of the user. Each emotion is assigned a number, meaning music and emotion are detected respectively [21].

V. Advantages of Proposed System

- The current model uses the Support Vector Machine (SVM) algorithm for classification, while we use Convolutional Neural Network (CNN), which is higher and better than SVM [22].
- One of the most unique features of CNN is its ability to identify the most important features in an image without human assistance. They are non-linear and are designed to recognize patterns and features in images with high accuracy.
- The layers used in CNNs have proven to be more efficient because the more layers we add, the higher the model competition and better results. Each image is converted to an array of pixels and SVM is different.

VI. Results

After image classification, we can process and analyse people's emotions. The different emotions are anger, sadness, and happiness.

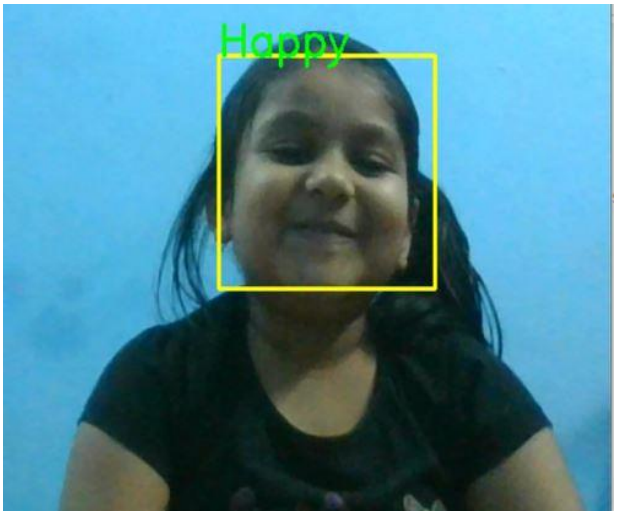

Fig. 7A. Happy face generated according to the emotion (happy)

```
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 33ms/step
Happy
['Paatshala', 'Saat Samunder Paar', 'Chikni Chameli', 'Chhookar Mere Man Ko Kiya Toone Kya Ishaara',
mmy', 'Main Shayar To Nahin', 'Let's Nacho', 'Baby Ko Bass Pasand Hai', 'Mera Joota Hai Japani']
```

Fig. 7B. Generated playlist for happy songs

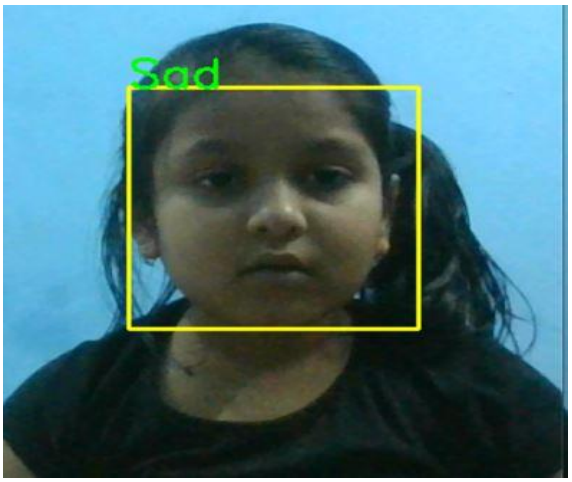

Fig. 8A. Sad face as per the emotion sad

```
1/1 [==============================] - 0s 32ms/step
1/1 [==============================] - 0s 48ms/step
1/1 [==============================] - 0s 40ms/step
1/1 [==============================] - 0s 32ms/step
1/1 [==============================] - 0s 46ms/step
1/1 [==============================] - 0s 45ms/step
1/1 [==============================] - 0s 41ms/step
1/1 [==============================] - 0s 34ms/step
Sad
["Sapna Mera Toota']", "Maine Mere Jaana']", "Mohabbat Ki Raahon Mein']", "Jaane Kyun
a Aaja Ve ']", "O Meri Mehbooba']", "Main Dardo Ko']", "Sapna Mera Toota']", "Kabira']
```

Fig. 8B. Generated playlist for sad songs

Fig. 9A. Angry face depicting emotion anger

```
1/1 [==============================] - 0s 40ms/step
1/1 [==============================] - 0s 38ms/step
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 38ms/step
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 37ms/step
Angry
["Bulleya (Ae Dil Hain Mushkil)']", "Tumhe Apna Banena ka (Hate Story 3)']"
a Aapne)']", "Galliyaan (Ek Villain)']", "Patakha guddi']", "Ghani Bawri']"
Tu Lucky Me (Humpty Sharma ki Dulhaniya)']"]
```

Fig. 9B. Generated playlist for angry songs

Several research papers have been accessed, that employed support vector machine (SVM), extreme learning machine (ELM), and convolutional neural network. The below table demonstrates the contrast of similar algorithms. The accuracy and algorithms that correspond. Each study has been assigned values. Convolutional neural networks enhance the precision of emotion detection by increasing its effectiveness [23].

TABLE I. ACCURACY OF VALIDATION AND TESTING FOR THREE DIFFERENT ALGORITHMS ON THE FER-2013 DATASET

Algorithm used	**Support Vector Machine**	**Extreme Learning Machines**	**Convolutional Neural Network**
Accuracy of Validation	68.67	65.76	71.10
Accuracy of Testing	69.45	64.89	91.68

VII. CONCLUSION

In this project, the music recommendation model is based on real-time feedback from users. This project is designed to help create a better relationship between music and users. Because music helps change the mood of the wearer and relieves stress for some. Recent developments hold great promise for creating music based on emotions. Therefore, this system recognizes our facial expression in different behaviors and plays music accordingly. In this system, we offer guidance on how to choose the greatest music tracks to uplift a user's spirits as well as a general explanation of how music might affect a user's mood. The system in place can detect user emotions. The current system has the capacity to recognize user emotions. The system could distinguish between happy, sad, angry, neutral, and surprised feelings. The suggested method asked the user to identify their feelings and then provided a playlist of songs that matched that emotion. Processing a huge dataset result in an increase in memory and CPU use. As a result, development will be harder and more interesting. The objective is to create this application on a common platform as economically as possible.

REFERENCES

[1] P. Rajashree, M. Sahana, and H. Savitri, “Review on facial expression-based music player,” International Journal of Engineering Research & Technology (IJERT), ISSN 2278-0181, vol. 6, no. 15, 2018.

[2] D. Reney, and N. Tripaathi, “An Efficient Method to Face and Emotion Detection,” Proceedings of Fifth International Conference on Communication Systems and Network Technologies, 2019.

[3] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang, “A survey of affect recognition methods Audio, visual, and spontaneous expressions IEEE transactions on pattern analysis and machine intelligence, vol. 31, pp. 39-58, 2008.

[4] K. Shantha Shalini, R. Jaichandran, S. Leelavathy, R. Raviraghul, J. Ranjitha, and N. Saravanakumar, “Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques,” Turkish Journal of Computer and Mathematics Education, vol. 12, no. 2, pp. 912–917, 2021.

[5] R. Ravi, S. V. Yadhukrishna, and R. prithviraj, "A Face Expression Recognition Using CNN & LBP," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 684-689, 2020.

[6] T. Shen, J. Jia, Y. Li, Y. Ma, Y. Bu, H. Wang, B. Chen, T. S. Chua, and W. Hall, "Peia: Personality and emotion integrated attentive model for music recommendation on social media platforms", Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 1, pp. 206-213, 2020.

[7] D. Gossi, and M. H. Gunes, “Lyric-based music recommendation,” in Studies in computational intelligence, Springer Nature, pp. 301–310, 2020.

[8] B. Shao, D. Wang, T. Li, and M. Ogihara, “Music recommendation based on acoustic features and user access patterns,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 8, 2009.

[9] J. Zhang, “Movies and Pop Songs Recommendation System by Emotion Detection through Facial Recognition,” Journal of Physics: Conference Series – IOPscience, vol. 1650, pp. 032076, 2020.

[10] D. Ayata, Y. Yaslan, and M. E. Kamasak, “Emotion based music recommendation system using wearable physiological sensors,” IEEE Transactions on Consumer Electronics, vol. 64, no. 2, pp. 196–203, 2018.

[11] S. Planet, and I. Iriondo, “Comparison between decision-level and feature-level fusion of acoustic and linguistic features for spontaneous emotion recognition,” in Proceedings of Iberian Conference of Information Systems and Technologies., Madrid, Spain: IEEE, 2019.

[12] S. Swaminathan, and E. G. Schellenberg, “Current emotion research in music psychology,” Emotion Review, vol. 7, no. 2, pp. 189–197, 2022.

[13] V. K. Gupta, A. Gupta, P. Jain, and P. Kumar, “Linear B-cell Epitopes Prediction using Bagging based Proposed Ensemble Model,” International Journal of Information Technology, Springer Nature, vol. 14, pp. 3517–3526, 2022.

[14] V. K. Gupta, “Toxicity Detection of Small Drug Molecules of the Mito -chondrial Membrane Potential Signalling Pathway,” International Journal of Data Mining and Bioinformatics”, vol. 27, no. 1/2/3, 2022.

[15] P. Kumar, V. K. Gupta, and D. P. Singh, "Face Mask Detection Using Convolution Neural Network," 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 2022.

[16] A. Tomar, S. Kumar, B.Pant, and U. K. Tiwari, “Dynamic Kernel CNN-LR model for people counting,” Applied Intelligence, vol. 52, pp. 1-16, 2022.

[17] V. Tripathi, A. Mittal, D. Gangodkar, and V. Kanth, “Real time security framework for detecting abnormal events at ATM installations,” Journal of Real-time image processing, vol. 16, pp. 535-545, 2019.

[18] P. Madan, V. Singh, D.P. Singh, M. Diwakar, B. Pant, and A. Kishor, “A Hybrid Deep Learning Approach for ECG-Based Arrhythmia Classification,” *Bioengineering*, vol. 9, issue 04, pp. 152, 2022.

[19] S. K. Mishra, A. Gupta, A. Tomar and V. K. Gupta, "Health Care Prediction for Various Diseases using Computational Intelligence Approaches: A Review," 2023 World Conference on Communication & Computing (WCONF), Raipur, India, pp. 1-6, 2023.

[20] M. K. Singh, P. Singh, V. K. Gupta, K. Mishra and A. Gupta, "Performance Analysis of CNN Models with Data Augmentation in Rice Diseases," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-5, 2023.

[21] N. Jaiswal, V. K. Gupta and A. Mishra, "Survey paper on various techniques of recognition and tracking," 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, pp. 921-925, 2015.

[22] V. K. Gupta, A. Gupta, M. K. Singh, A. Gupta and A. Tomar, "Toxicity Detection Methodology of Adverse Outcome Pathways using Physicochemical Properties and Machine Learning Approaches," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-7, 2023.

[23] V. K. Gupta, A. Gupta, V. Tyagi, P. Pandey, R. Gupta and D. Kumar, "Multilevel Face Mask Detection System using Ensemble based Convolution Neural Network," 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, pp. 391-396, 2023.