ISSN:1533-9211
SEYBOLD REPORT
Journal
www.seybold-report.com

Predicting Human Personality using Multimedia by Employing Machine Learning Technique

Akshata S Bhayyar*[1], Kiran P[2], Sushma Bylaiah[1]

[1]Assistant Professor, Ramaiah Institute of Technology, Bengaluru, India

[2]Professor, RNS Institute of Technology, Bengaluru, India

ISSN:1533-9211
SEYBOLD REPORT
Journal
www.seybold-report.com

ISSN: 1533 - 9211

Abstract

Recently, cognitive-based sentiment analysis has drawn a lot of attention because it focuses on automatically identifying user behaviours like personality characteristics from online social media text. In order to demonstrate the effectiveness of the suggested model for eight key personality traits (Introversion-Extroversion, Intuition-Sensing, Thinking-Feeling, and Judging-Perceiving), we present a hybrid Deep Learning-based model made up of Convolutional Neural Networks with Long Short-Term Memory. On the basis of audio and video recordings of human faces, we provide a model for the identification of personality traits. A web-based platform is created to gather the dataset, allowing users to record voice and video using a microphone and webcam, respectively. The dataset contains videos and audio clips of people of various ages and genders. Applying the proposed CNN+LSTM model on the considered dataset we could achieve an accuracy of 87.07%.

CORRESPONDING AUTHOR:

Akshata S Bhayyar

akshataphd@gmail.com

KEYWORDS:

Cognitive Psychology, Personality traits, Convolutional Neural Networks, Long Short-Term Memory, Myers-Briggs Type Indicator (MBTI)

Received: 22 October 2023
Accepted: 05 November 2023
Published: 13 November 2023

TO CITE THIS ARTICLE:

Bhayyar, A. S., Kiran , P., & Bylaiah, S. (2023). Predicting Human Personality using Multimedia by Employing Machine Learning Technique. *Seybold Report Journal*, *18*(09), 102-120. DOI:10-5110-77-1046

1. Introduction

The amount of data on the Web is growing tremendously every day, practically every second. The majority of this information—text, audio, and video—is supplied by internet users who are increasingly sharing content on social media, blogs, online discussion forums, and other websites with a comparable audience. The phrase "big data" is often used to describe extraordinarily huge datasets because to the exponential growth of multimodal data. Contrary to standard datasets, big data often consists of enormous amounts of unstructured data that need real- time analysis [1]. The study of personality recognition is an emerging field in the current era. A person's behavior, mood, motivation, and thought patterns are all parts of their personality. Our personalities have a significant impact on our lives, influencing our decisions in terms of our goals, ambitions, and mental and physical health. Automatic personality prediction can therefore be used for a variety of purposes, such as enhanced personal assistants, recommender systems, job screening, forensics, psychological studies, political predictions, specialised therapy, etc. [2].

The five personality qualities Extraversion, Neuroticism, Agreability, Conscientiousness, and Openness are the most often used metrics in the literature. Extraversion provides details about a person's gregariousness, excitement, assertiveness, talkativeness, and high level of emotional expressiveness. Instability, sadness, and emotional instability are all signs of neuroticism in an anxious individual. Being agreeable reveals a person's generosity, modesty, complexity and unreliability. The personality traits of the user can be examined using image, audio, and video samples. The extraction of features from images—most notably the face features—has been done using CNN-based architectures. Many sources, including OpenSMILE, have been used to get the audio attributes, including the MelFrequency Cepstral Coefficients (MFCC), Logfbank, Zero Crossing Rate (ZCR), pitch, and loudness. In order to combine the results produced by many modalities, researchers typically employ the late fusion technique such as averaging the predictions from all phases[3]. Psychologists have long examined human personality, and numerous ideas have been put out over time to classify, explain, and comprehend it. Models based on attributes are the most accurate in forecasting quantifiable aspects of a person's life. Determining and assessment of traits—habitual patterns of behaviours, thoughts, and emotions that are generally stable across time—are the foundation of the approach known as trait theory. The foundation of trait models is in human evaluations of the semantic similarities and connections between the adjectives that people use to describe one another and themselves[4]. An individual's sentiment can be analysed in a variety of ways, including through the study of facial expressions, audio and video recordings, real-time conversation design, etc[5]. Deep learning models like Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) [6] have achieved recent progress in action recognition. Convolutional Neural Network (CNN) is a sophisticated artificial network that, before being linked to further image segmentation algorithms, detects visual forms on input images with minimal processing. It has four convolutional layers, four compound layers to extract the elements, two fully linked layers, and a softmax layer with seven sensory classes. [7].

Given that the majority of individuals are struggling with stress issues, the intensity of human emotions has a big impact on one's sense of security and current condition. The implications for related fields like social robotics, cognitive science, and behavioral sciences will be greater if a model can recognize a person's facial expressions, decode their emotions, and open up a new area of research into the identification of socio-emotional phenomena (affect, personality, mental health, and engagement) from simulations of person-specific cognitive processes.

2. Proposed methodology

The architecture of the model consists of four steps, each of which is applied differently to text data , audio, and video data. Initially, the data is acquired in different forms viz., text, audio, and video. The proposed architecture comprising all four stages as seen in Figure 1.

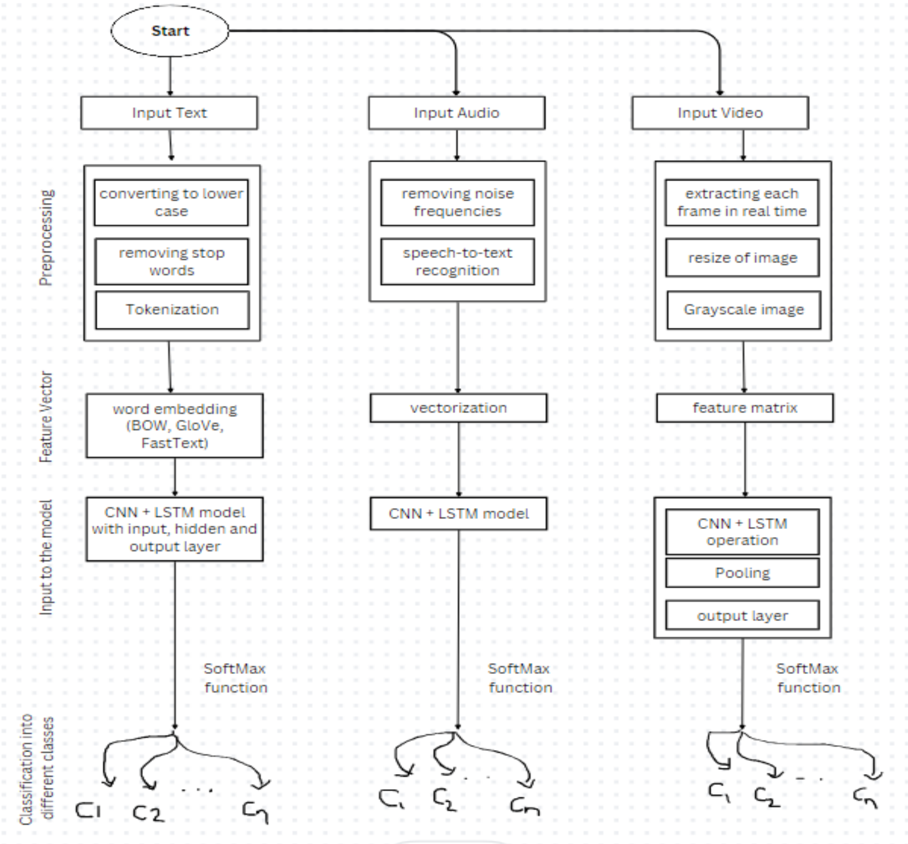

Figure 1. Architecture of the proposed model

2.1 Data collection

Using HTML, CSS, and Javascript, we created a website that would enable people to videotape and audio record themselves (Figure 2). The participant believes that using the internet as a tool to record videos is the most practical and accessible option because it functions flawlessly across all operating systems [8]. To start the recording process, the website requests that visitors access the camera on their device. The participant can start and stop the recording process. With this approach, we encountered the issue of several video formats that the participants had recorded in, as a result, we turned every film into a common format, i.e .webm, which aided for preprocessing.

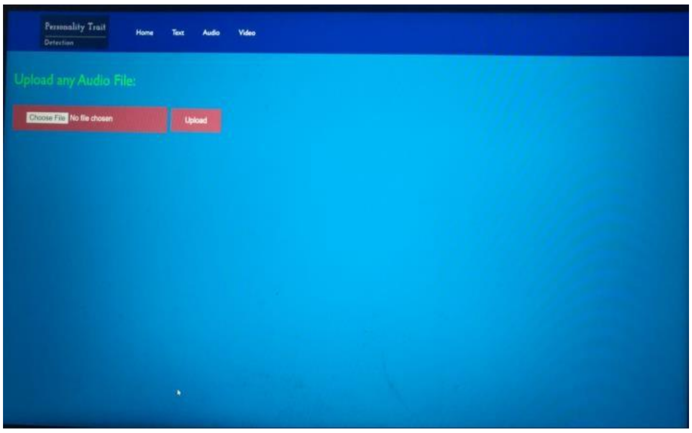

Figure 2. Screenshot of website

2.2 Data Preprocessing

2.2.1 Text Preprocessing

For NLP (natural language processing) activities, the largest difficulty is the discrepancy between social media text data and conventional english. Prior to feature extraction, all social media data sets need to be pre-processed. Because the Facebook datasets are written in English and the Twitter dataset in Bahasa, pre-processing normalises the text datasets. [9]. The data mining process includes data pre-processing. Real-world data is not limited to a single field and is gathered in a number of ways, leading to erroneous, insufficient, and unreliable data. Various techniques are employed in our framework throughout the pre-processing stage. The initial stage eliminates the user-defined text patterns. Patterns such as "user handles (@username)", "hashtags (#hashtag)", "URLs", "characters, symbols, and numbers other than alphabets", "empty strings", "drop rows with NaN in the column", "duplicate rows", etc. are eliminated [10] [11].

In the second stage stopwords, that don't add to the content of the sentence including "is," "was," "at," "if," and similar words are removed, and stemming is carried out. To eliminate stopwords from our text, the NLTK package, which includes a collection of stopwords, is employed. A word can be changed to its basic form via stemming. By deleting prefixes or suffixes (such as "ize," "ed," "s," "de," etc.) from a word, Porter Stemmer is used to construct the term's root. The cleaned tweets are then returned and given to the tokenizer as input after all of the tweets have been cleaned. One of the primary pre-processing steps for NLP systems is tokenizing raw text input. Tokenizers are instruments that divide a given string using regular expressions [12]. The subsequent phase in the tokenization procedure is the texts_to_sequences() technique. It accepts data from the prior technique made up of as many words with an index as possible. The program's goal is to convert each word in a tweet into a string of integers before replacing each one with its corresponding integer value from the word_index vocabulary. The tweets are now been changed into various length of integer groupings.

2.2.2 Audio Preprocessing

Vocal features can be precisely determined from a remarkably small amount of audio waveform data. Additionally, the linguistic features in the current study directly make use of sample length [13]. Using the Google Speech-to- text API, we first processed audio files from a user and converted them into a transcript [14]. The system combines two two-pass english systems that respectively use acoustic models based on single-distant microphones (SDM) and individual head-mounted microphones (IHM). The common decoding configurations are used in the IHM and SDM systems. For example, the first pass employs unadopted acoustic models while the second pass uses unsupervised adaptation, where acoustic models trained with SAT(speaker adaptive training) are less dependent on training speakers and perform better when applied to test speakers that are not visible [15].

2.2.3 **Video preprocessing**

The video is captured using OpenCV python module through the webcam of the device. It loops continuously to capture and display frames from the video. It resizes the frame at a size of 64X64 and converts the frame to a image and saves it in png format and focuses on the features of the face such as eyes, nose and mouth as in Figure

3, The model loads an image from a specified directory using the OpenCv Python module [16]. Training and Testing of the image is done by keras ImageDataGenerator. The original data is fed into the Keras ImageDataGenerator, which transforms it at random and produces a result that solely contains the newly changed data. Using the picture data generator offered by Keras, we may loop through the data in batches. The field of real-time data augmentation employs the Keras image data generator to produce batches that comprise the information from tensor images. [17].

2.3 Feature Extraction

2.3.1 **Text Extraction**

The first step when working with text is to devise a method for converting strings to numbers (or to "vectorize" the text) before alimentation it to the model as vectors (arrays of numbers) are the input to machine learning models. Bag of Words (BOW) algorithm is used for word embedding. The portrayal of words with similar meanings is made possible via word embeddings, a type of word representation. They are a distributed representation of text, which may be a crucial discovery in enabling deep learning algorithms to perform so well on difficult natural language processing problems. [17]. It records how frequently a word or phrase appears in a document. For uses like search, document categorization, and topic modelling, those word counts enable us to compare documents and evaluate their similarity. BOW is a technique for getting text ready for deep learning networks as in Figure 3. A collection of text is transformed into a matrix of token counts using CountVectorizer. It extracts one word from each text, ignoring stop words because counting them makes no sense. It first counts the words and then creates a Sparse matrix. For static text word embedding techniques like Glove and fast text can be used.

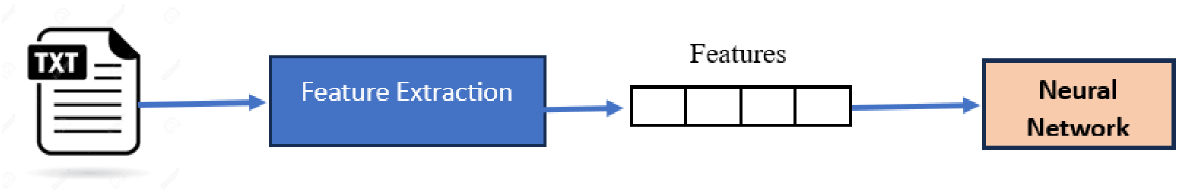

Figure 3. Text feature extraction

2.3.2 Audio/video feature extraction

The prosodic (i.e., nonverbal) and acoustic elements of speech are examined on the audio channel. It allows raw waveforms as input instead of frequently used spectrograms or traditional feature sets because CNNs can automatically extract relevant properties from audio. [17].The input images' three channels (blue, red, and green) are supplied into the model after the video has been converted to images. Following are a number of convolutional layers paired with padding and max-pooling layers. To map the personality qualities, we employ two final completely connected layers with sigmoid activation. Only these two layers are used for training. Pre-trained models can often be fine-tuned to take advantage of training on huge external datasets (Figure 4).

Figure 4: Extraction of features from video using CNN model

2.4 CNN + LSTM Model

To categorize the input transcript into the cognitive psychology of personality traits, the anticipated CNN-LSTM model functioned in the subsequent layers. The embedding layer serves as the input layer, the CNN serves as the hidden layer, and the dense layer as the output layer with the Softmax activation function. [18][19][20].

2.4.1 Embedding layer

In the anticipated CNN-LSTM model, the embedding layer is the first layer. Given an input review of n words, it transmutes each word into a real-valued vector, for example, where the variable specifies the word's dimension. A feature matrix is consequently produced that shows how long the input review took. After that, the CNN layer receives the result as input.

2.4.2 Convolutional Neural Network Layer

CNNs are great feature extractors because they can retrieve both high-level qualities like edges and low-level properties like objects in images during each epoch [21][22]. Time dependencies and particular features can also be captured by using the right filters, convolution layers, reduction of dimensionality operation, and pooling approaches like maximum pooling and Average-pooling. By compressing the spatial dimension of the convolution matrices into a form that is simpler to process without relinquishing critical classification information, a convolution layer affects the input properties. Kernel or Filter is the name given to the matrix used to realise convolution information [18]. The kernel travels with a specific stride whenever a multiplication of a matrix operation occurs between the kernel and the part of the input vector that supersedes itself on the kernel. This reduces the amount of processing resources needed to handle the input while also assisting in the identification of significant traits that are location invariant and subsidized to operative model training. Convolutional layers, combining layers, drop-out layers, and dense layers have all been applied in the present work. Convolutional layers were mostly utilized by the neural network during training [5], [23].

2.4.3 Long Short-Term Memory

Long-short-term memory (LSTM) refers to an artificial recurrent neural network (RNN) with long-range dependencies [24]. The output layer, known as Softmax, does additional processing after the LSTM layer creates a feature vector using input from the pooling layer (Figure 5).

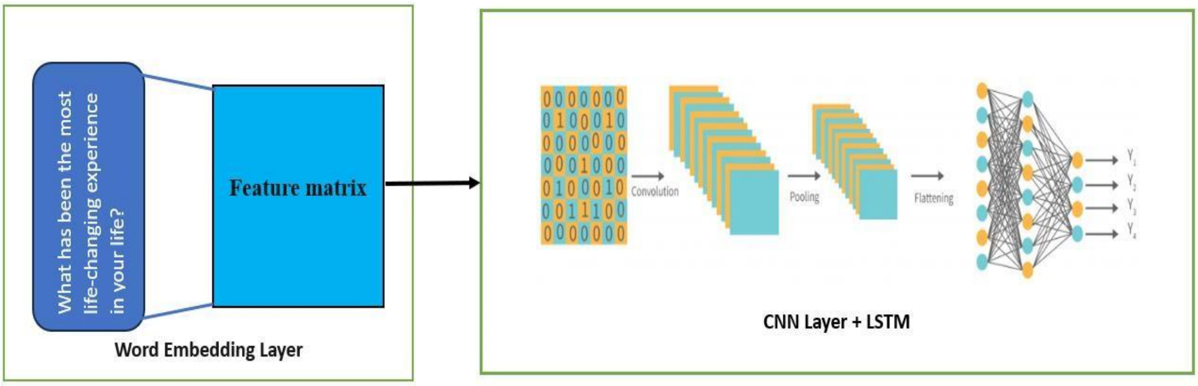

Figure 5. CNN + LSTM model

With the help of ACCURACY, the proposed model's discriminating ability was assessed. The model that met the requirements with ACCURACY >0.80 was considered to make the best forecast[24][25].

3. Results

3.1 Data Preprocessing

3.1.1 Text Preprocessing

The outcome of the proposed work for the majority and minority classes is displayed in Table 1. The greatest F1- score value achieved is 0.49, and the validation's precision is 0.55.

Table 1. LSTM Training

Personality Traits	Precision	Recall	F1-score	Support
ISTJ	0.07	0.05	0.06	43
ISTP	0.14	0.12	0.13	139
ISFJ	0.07	0.26	0.11	35
ISFP	0.31	0.23	0.26	138
INFJ	0.00	0.00	0.00	13
INFP	0.00	0.00	0.00	8
INTJ	0.00	0.00	0.00	8
INTP	0.00	0.00	0.00	16
ESTP	0.44	0.48	0.46	295
ESTJ	0.55	0.50	0.52	356
ESFP	0.45	0.33	0.38	239
ESFJ	0.47	0.51	0.49	277
ENFP	0.09	0.06	0.07	35
ENFJ	0.19	0.26	0.22	47
ENTP	0.19	0.38	0.25	24
ENTJ	0.22	0.31	0.26	62
micro avg	0.37	0.37	0.37	1735
macro avg	0.20	0.22	0.20	1735
weighted avg	0.38	0.37	0.37	1735
samples avg	0.37	0.37	0.37	1735

The primary goal of the present research is to examine the impact of various fusion methods, audio, visual, and text modalities, on personality detection.

3.1.2 Audio Preprocessing

Figure 6 depicts the audio-to-text conversion by using Google Speech-to-text API. The speech algorithms on the Google Cloud convert audio to text using neural network models, and they may be used locally on any device.

```
Speak something...
result2:
{   'alternative': [   {   'confidence': 0.89401221,
                           'transcript': 'product is very good and everyone '
                                         'should buy it'},
                       {   'transcript': 'product is very good and everyone '
                                         'should buy'},
                       {   'transcript': 'product is very good and everyone '
                                         'should b'},
                       {   'transcript': 'product is very good and everyone '
                                         'should I buy it'},
                       {   'transcript': 'product is very good and everyone '
                                         'should buy this'}],
    'final': True}
You said: product is very good and everyone should buy it
product is very good and everyone should buy it
```

Figure 6. Converting audio into transcript using a Google to Speech-to-text

3.1.3 Video Preprocessing

The results (Figure 7) determine the importance of facial features, particularly the eyes, nose, and mouth, in predicting personality traits. In order to estimate a person's personality, the model proposed a multi-modal CNN for acquiring visual details from brief clips of video in combination with predictions of qualities specific to the face. [26], [27].

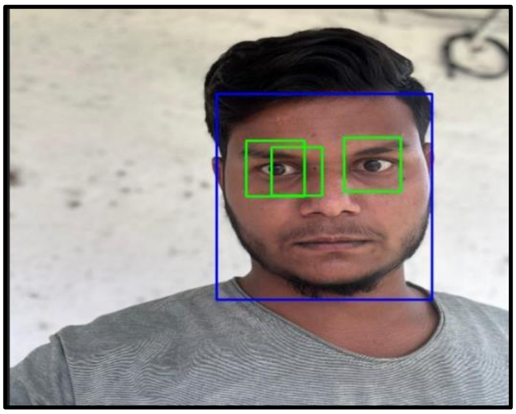

Figure 7. A video is converted to an image and features of the face are extracted

3.2 CNN + LSTM Model

3.2.1 **Audio results**

A number of psychologically significant phenomena can be expressed through the powerful medium of speech. The device's microphone records audio, which is then automatically converted to a transcript. The CNN+LSTM model extracts features from the transcript and predicts an individual's personality, as illustrated in figure 8a and 8b. The transcript in figure 8a, it analyses that the personality of a person is of type ENFJ figure 8b, ENFJs are particularly skilled at bringing disparate groups of people together. As a result, they may make great influence and permeate a group with a desire that is both inspirational and motivating.

```
Speak something...
result2:
{   'alternative': [   {   'confidence': 0.88687521,
                           'transcript': 'ntp are injured or conversation the '
                                         'other day is a territory gapping '
                                         'about the nature of the universe and '
                                         'the idea that every and social code '
                                         'b arbitrary construct created'},
                       {   'transcript': 'ntini are injured or conversation '
                                         'the other day is a territory gapping '
                                         'about the nature of the universe and '
                                         'the idea that every and social code '
                                         'b arbitrary construct created'},
                       {   'transcript': 'ntp are injured or conversation the '
                                         'other day is a territory gapping '
                                         'about the nature of the universe and '
                                         'the idea that Avery and social code '
                                         'b arbitrary construct created'},
                       {   'transcript': 'ntp are injured or conversation the '
                                         'other day is a territory gapping '
                                         'about the nature of the universe and '
                                         'the idea that every and social code '
                                         'b arbitrary construkt created'},
                       {   'transcript': 'ntp are injured or conversation the '
                                         'other day is a territory gapping '
                                         'about the nature of the Universe end '
                                         'the idea that every and social code '
                                         'b arbitrary construct created'}],
    'final': True}
You said: ntp are injured or conversation the other day is a territory gapping about the nature of the universe and the idea th
at every and social code b arbitrary construct created
ntp are injured or conversation the other day is a territory gapping about the nature of the universe and the idea that every a
nd social code b arbitrary construct created
```

Figure 8a. Audio is converted to a transcript

```
The predicted trait from the given audio is: ENFJ - Extroversion:Intuition:Feeling:Judging
```

Figure 8b. Using The CNN+LSTM model the personality is predicted from the transcript

3.2.2 **Video Results**

Functionals summarise the facial traits that were retrieved from a whole video segment. The only part of each video that is used to extract scene elements is the face image. It is believed that videos do not span several shots. Faces are recognized, positioned correctly, and scaled to 64 X 64 pixels. In a neural network using video face training, video pictures are summarised by calculating the personality of a person in the image after each aligned face's frame-level attributes are extracted. Figures 9a,9b,9c,9d represents the prediction of the personality of an individual. The description of Figure 8 is illustrated in the Table 2.

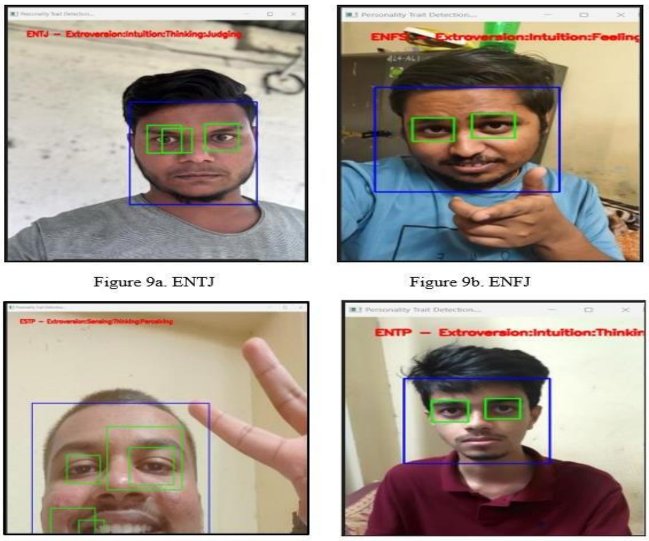

Figure 9a. ENTJ Figure 9b. ENFJ

Figure 9c. ESTP Figure 9d. ENTP

Figure 9. Personality traits identified by CNN+LSTM model

Table 2. Description of personality traits

Figure No	Personality Type	Description
8a	ENTJ	Extraversion: Intuitive: Thinking: Judging
8b	ENFJ	Extraversion: Intuitive: Feeling: Judging
8c	ESTP	Extraversion: Sensing: Thinking: Perceiving
8d	ENTP	Extraversion: Intuitive: Thinking: Perceiving

4. Discussion

The research aims to investigate how CNN-LSTM affects the personality prediction of social network users and to increase the model's accuracy for such tasks [28].To analyse the personality of the person based on text/audio/video is trained using an MBTI dataset [29] consisting of personality traits, which consists of 8.675 rows of forum posts from PersonalityCafe with labels for the 16 MBTI personality types of their authors. Sensation, intuition, feeling, and thinking are the four main psychological functions that are classified using the MBTI framework. Based on Jung's explanation of personality as shown in Table 3 [30], [31], the MBTI is a psychological tool that contains assessments for attitude and functions. Users can better understand their own personality traits with the MBTI, particularly their preferences, dislikes, strengths, and weaknesses, as well as possible career proclivities and compatibility with others.

Table 3. The Myers-Briggs technique's four dichotomies

Type	Description
Extraversion (E) – Introversion (I)	Extraverts, also known as extroverts, are "outward-turning" and tend to be action-oriented. They like frequent social interaction and report feeling revitalised afterward. Introverts are "inward-turning" and tend to be thought-oriented. They also value deep and meaningful relationships with others and find that time spent alone themselves is reviving.
Sensing (S) – Intuition (N)	People who enjoy perceiving are typically astute observers of reality, particularly what their senses can show. They tend to seek out personal knowledge and concentrate on specific details and facts. People who value their intuition highly are more prone to focus on patterns and impressions. They enjoy peering into the future, making predictions, and thinking across ideas.
Thinking (T) – Feeling (F)	People who want to think more highly of facts and objective knowledge. They frequently demonstrate consistency, logic, and

	impartiality when making decisions. Those who value sensation are more prone to consider people and their emotions when making decisions.
Judging (J) – Perceiving (P)	People with a propensity to judge long for structure and authority. They are more perceptive-inclined more nimble, adaptable, and understanding.

In the beginning, we trained a single 16-class classifier to test whether our deep network could produce a better outcome than our baseline softmax model [30][31], [32]. Table 4 shows the details of all 16-personality traits.

Table 4. The 16 different Myers-Briggs Type Indicator (MBTI) personality types

Personality code	**Type of Person**	**Description**
ISTJ	The Inspector	They are typically loyal, organized, conventional, reserved, and sensible.
ISTP	The Crafter	They value opportunities for hands-on learning and are fiercely independent.
ISFJ	The Protector	Because they are sympathetic and dedicated, they are always ready to stance up for the people they care about.
ISFP	The Artist	They tend to be quiet, artistic, adaptable, and laid-back.
INFJ	The Advocate	They are imaginative and analytical, and are thought to be one of the scarcest Myers-Briggs types.
INFP	The Mediator	They have high moral standards and aim to make the world a better place.
INTJ	The Architect	They are equally inventive and analytical, and they have good logic.
INTP	The Thinker	They are renowned for having a ironic inner world and for being low and introverted.
ESTP	The Persuader	They enjoy engaging with others, are extroverted and dramatic, and focus on the here and now.
ESTJ	The Director	They tend to take on leadership duties, are forceful, and adhere to the rules.
ESFP	The Performer	They are social and like to be the centre of attention.
ESFJ	The Caregiver	They enjoy thinking positively about others and have an open demeanor.
ENFP	The Champion	Being charismatic and effervescent, they enjoy circumstances where they may employ their imagination.
ENFJ	The Giver	They are known as being devoted and sensitive, as well as understanding and caring.
ENTP	The Debater	They appreciate being surrounded by ideas, are very creative, and frequently start numerous projects at once.
ENTJ	The Commander	Because they are aggressive and self-assured, they make good project planners.

In terms of neural network configuration [20][33][34], We used an LSTM network with the subsequent specifications: 100 epochs, fixed input length, two layers, learning rate of 0.001, hidden layer dimension of 1500, embedding dimension of 200, dropout of 0.2, validation phase dataset of 30%, maximum terminology size of 25.000, batch size of 32, and ADAM as an optimizer. We cast-off Cross-entropy loss as a loss function when training neural networks.
The model is trained with 76.9Mb of personality trait labelled dataset as in Figure 10 shows the samples of the image's dataset. The dataset is used for both image and video distinguishing personality traits as video is converted to the frame of images [16] [4][5] to understand the

personality of an individual.

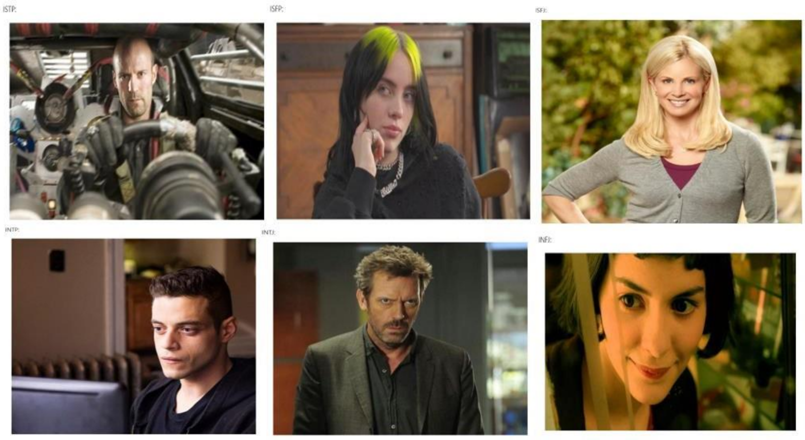

Figure 10. Images of the data set's samples. The image from top left to bottom right are labeled as ISTP, ISFP, ISFJ, INTP, INTJ, INFJ

Comparing our method to the earlier models, it is observed that CNN+LSTM consistently improves stability and performance. The personality trait accuracy metric achieved is 0.8707 Table 5.

Table 5. Analysis of recent studies with text/audio/video data with our proposed work

Ref	**Author and Year**	**Dataset used**	**Type of Data**	**Technique used**	**Accuracy**
[35]	Liangqing Wu, Dong Zhang et al. 2020	There are 1,000 video movie reviews in the Persuasive Opinion Multimedia (POM) corpus collection.	Text, Audio, Video	Multimodal explicit many2many interactions (MEMI)	The highest value of 47.3% for the Lazy category, and the rest are below 47.3%
[36]	Liu Z, Rehman A et al. 2021	SSPNet Speaker Personality Corpus	Audio	SAM(split-at-mean) and BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)	SAM - 76.9% BIRCH- 76.4%
[37]	Song S, Jaiswal S et al. 2023	Video-Based Self-Reported Personality Prediction	Video-based	Multi-scale models	85.40%
[6]	Beyan C, Zunino A et al. 2019	ELEA-AV Corpus, ChaLearn First Impression Dataset	Images	CNN +LSTM	77%
[5]	Singh S, Kumar A et al. 2021	FER2013 dataset	Images	CNN	86%
Proposed Model	**Akshata S Bhayyar et al. 2023**	**Text and Audio, Video real-time recording**	**Text, Audio, Video**	**CNN + LSTM**	**87.07%**

5. Conclusion

The present work combines CNN+ LSTM architecture to identify true personality from person-specific human cognition. Experimental findings are based on text, audio, and video data. The audio data is converted to a transcript and video data is converted to an image detecting an individual's facial expression to analyse a person's personality among the 16 personality traits by Myers Briggs. The CNN+LSTM model achieves an accuracy of 87.07%, which is a better result in comparison with the existing literature work. As far as we know, this is the first work on human personality prediction considering 16 personality traits using multimedia. This research thus opens a new line of research into the identification of socio-emotional phenomena (personality, engagement, mental health, and affect) from imitations of person-specific cognitive processes, with further insinuations for pertinent fields like social robotics, behavioral sciences and cognition.

COMPETING INTERESTS

The authors have no competing interests to declare.

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Data availability statement

The Myers-Briggs Personality Type dataset was used for the research work. Link for accessing the dataset https://www.kaggle.com/datasets/datasnaek/mbti-type.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author's Affiliation

Akshata S Bhayyar*[1], Kiran P[2], Sushma Bylaiah[1]
akshataphd@gmail.com, kiranpmys@gmail.com, sushma.b3@gmail.com
[1]Assistant Professor, Ramaiah Institute of Technology, Bengaluru, India
[2]Professor, RNS Institute of Technology, Bengaluru, India

*Corresponding Author: Akshata S Bhayyar, Email Id: akshataphd@gmail.com

Author's Profile

Mrs. Akshata S. Bhayyar, currently working as Assistant Professor, Dept. of CSE, RIT. Having Teaching experience of 11 years and areas of interest are data mining, sentimental analysis, cognitive psychology in online social network

Dr. Kiran P, BE, MTech, Ph.D., working as HoD, Professor in CSE Department, RNSIT, Bengaluru from past 24 years. His research interests include Cryptography, Big Data Analytics, Data Mining, Privacy Preserving Data Mining and Privacy Preserving Data Publishing.

Dr. Sushma B. is working as an Assistant Professor in Computer Science Department of Ramaiah Institute of Technology. Her areas of interest include Data mining, Machine Learning, Data Structures, and Database Systems. She has teaching experience of 13 years.

HOW TO CITE THIS ARTICLE:

Bhayyar, A. S., Kiran , P., & Bylaiah, S. (2023). Predicting Human Personality using Multimedia by Employing Machine Learning Technique. *Seybold Report Journal*, *18*(09), 102-120. DOI:10-5110-77-1046

References

[1] N. J. Shoumy, L. M. Ang, K. P. Seng, D. M. M. Rahaman, and T. Zia, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," *Journal of Network and Computer Applications*, vol. 149. Academic Press, Jan. 01, 2020. doi: 10.1016/j.jnca.2019.102447.

[2] C. Suman, S. Saha, A. Gupta, S. K. Pandey, and P. Bhattacharyya, "A multi-modal personality prediction system," *Knowl Based Syst*, vol. 236, Jan. 2022, doi: 10.1016/j.knosys.2021.107715.

[3] C. M. Chen and C. H. Wu, "Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance," *Comput Educ*, vol. 80, pp. 108–121, Jan. 2015, doi: 10.1016/j.compedu.2014.08.015.

[4] J. C. S. Jacques Junior *et al.*, "First Impressions: A Survey on Vision-Based Apparent Personality Trait Analysis," *IEEE Trans Affect Comput*, vol. 13, no. 1, pp. 75–95, 2022, doi: 10.1109/TAFFC.2019.2930058.

[5] S. Singh, A. Kumar, and S. Thenmalar, "Facial Emotion Analysis and Recommendation Using CNN," in *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1084–1090. doi: 10.1109/I-SMAC52330.2021.9640904.

[6] C. Beyan, A. Zunino, M. Shahid, and V. Murino, "Personality Traits Classification Using Deep Visual Activity-Based Nonverbal Features of Key-Dynamic Images," *IEEE Trans Affect Comput*, vol. 12, no. 4, pp. 1084–1099, 2021, doi: 10.1109/TAFFC.2019.2944614.

[7] Y. Li *et al.*, "CR-Net: A Deep Classification-Regression Network for Multimodal Apparent Personality

Analysis," *Int J Comput Vis*, vol. 128, no. 12, pp. 2763–2780, Dec. 2020, doi: 10.1007/s11263-020-01309-y.

[8] K. Kassab, A. Kashevnik, A. Mayatin, and D. Zubok, "VPTD: Human Face Video Dataset for Personality Traits Detection," *Data (Basel)*, vol. 8, no. 7, p. 113, Jun. 2023, doi: 10.3390/data8070113.

[9] M. D. Kamalesh and B. B, "Personality prediction model for social media using machine learning Technique," *Computers and Electrical Engineering*, vol. 100, May 2022, doi: 10.1016/j.compeleceng.2022.107852.

[10] P. Dataset, P. William, and A. Badholia, "3400 P William, Dr. Abhishek Badholia Evaluating Efficacy of Classification Algorithms on Personality Evaluating Efficacy of Classification Algorithms on Personality Prediction Dataset," vol. 19, no. 4, pp. 3400–3413, 2020, doi: 10.17051/ilkonline.2020.04.764728.

[11] G. D. Salsabila and E. B. Setiawan, "Semantic Approach for Big Five Personality Prediction on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 680–687, Aug. 2021, doi: 10.29207/resti.v5i4.3197.

[12] Universitas Gadjah Mada, Universitas Gadjah Mada. Departemen Ilmu Komputer dan Elektronika, Institut Teknologi Bandung, Institute of Electrical and Electronics Engineers. Indonesia Section, and Institute of Electrical and Electronics Engineers, *Proceedings of 2015 International Conference on Data and Software Engineering (ICODSE 2015) : Universitas Gadjah Mada, Indonesia, November 25th-26th, 2015.*

[13] Z. N. K. Marrero, S. D. Gosling, J. W. Pennebaker, and G. M. Harari, "Evaluating voice samples as a potential source of information about personality," *Acta Psychol (Amst)*, vol. 230, Oct. 2022, doi: 10.1016/j.actpsy.2022.103740.

[14] O. Kampman, E. J. Barezi, D. Bertero, and P. Fung, "Investigating Audio, Visual, and Text Fusion Methods for End-to-End Automatic Personality Prediction," May 2018, [Online]. Available: http://arxiv.org/abs/1805.00705

[15] Y. Miao, H. Zhang, and F. Metze, "Speaker Adaptive Training of Deep Neural Network Acoustic Models using I-vectors." [Online]. Available: http://www.cs.cmu.edu/

[16] H. J. Escalante *et al.*, "Modeling, Recognizing, and Explaining Apparent Personality from Videos," *IEEE Trans Affect Comput*, vol. 13, no. 2, pp. 894–911, 2022, doi: 10.1109/TAFFC.2020.2973984.

[17] O. Kampman, E. J. Barezi, D. Bertero, and P. Fung, "Investigating Audio, Visual, and Text Fusion Methods for End-to-End Automatic Personality Prediction," May 2018, [Online]. Available: http://arxiv.org/abs/1805.00705

[18] H. Kour and M. K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," *Multimed Tools Appl*, vol. 81, no. 17, pp. 23649–23685, Jul. 2022, doi: 10.1007/s11042-022-12648-y.

[19] F. M. Alotaibi, M. Z. Asghar, and S. Ahmad, "A Hybrid CNN-LSTM Model for Psychopathic Class Detection from Tweeter Users," *Cognit Comput*, vol. 13, no. 3, pp. 709–723, May 2021, doi: 10.1007/s12559-021-09836-7.

[20] L. Xiaoyan, R. C. Raga, and S. Xuemei, "GloVe-CNN-BiLSTM Model for Sentiment Analysis on Text Reviews," *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/7212366.

[21] S. Singh, A. Kumar, and S. Thenmalar, "Facial Emotion Analysis and Recommendation Using CNN," in *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1084–1090. doi: 10.1109/I-SMAC52330.2021.9640904.

[22] H. Y. Suen, K. E. Hung, and C. L. Lin, "Intelligent video interview agent used to predict communication skill

and perceived personality traits," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, Dec. 2020, doi: 10.1186/s13673-020-0208-3.

[23] G. Lu, Y. Liu, J. Wang, and H. Wu, "CNN-BiLSTM-Attention: A multi-label neural classifier for short texts with a small set of labels," *Inf Process Manag*, vol. 60, no. 3, 2023, doi: 10.1016/j.ipm.2023.103320.

[24] S. Bylaiah *et al.*, "Airplane Health Management System View project Disease Prediction Model to Assess the Impact of Changes in Precipitation Level on the Risk of Anthrax Infectiousness among the Livestock Hosts in Karnataka, India." [Online]. Available: https://www.researchgate.net/publication/358262477

[25] O. Allouche, A. Tsoar, and R. Kadmon, "Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS)," *Journal of Applied Ecology*, vol. 43, no. 6, pp. 1223–1232, Dec. 2006, doi: 10.1111/j.1365-2664.2006.01214.x.

[26] R. D. P. Principi, C. Palmero, J. C. S. J. Junior, and S. Escalera, "On the Effect of Observed Subject Biases in Apparent Personality Analysis from Audio-Visual Signals," *IEEE Trans Affect Comput*, vol. 12, no. 3, pp. 607–621, Jul. 2021, doi: 10.1109/TAFFC.2019.2956030.

[27] R. Harrington and D. A. Loffredo, "MBTI personality type and other factors that relate to preference for online versus face-to-face instruction," *Internet and Higher Education*, vol. 13, no. 1–2, pp. 89–95, Jan. 2010, doi: 10.1016/j.iheduc.2009.11.006.

[28] J. Zhao, D. Zeng, Y. Xiao, L. Che, and M. Wang, "User personality prediction based on topic preference and sentiment analysis using LSTM model," *Pattern Recognit Lett*, vol. 138, pp. 397–402, Oct. 2020, doi: 10.1016/j.patrec.2020.07.035.

[29] MIPRO Croatian Society and Institute of Electrical and Electronics Engineers., *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO) : proceedings, September 27-October 1, 2021, Opatija, Croatia*.

[30] B. Cui and C. Qi, "Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction."

[31] K. Orynbekova, A. Talasbek, A. Omar, A. Bogdanchikov, and S. Kadyrov, "MBTI personality classification using Apache Spark," in *Proceedings - 2021 16th International Conference on Electronics Computer and Computation, ICECCO 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICECCO53203.2021.9663858.

[32] S. Yenikent, "Why Is MBTI Personality Detection from Texts a Difficult Task?"

[33] Y. Bao, Z. Huang, L. Li, Y. Wang, and Y. Liu, "A BiLSTM-CNN model for predicting users' next locations based on geotagged social media," *International Journal of Geographical Information Science*, vol. 35, no. 4, pp. 639–660, 2021, doi: 10.1080/13658816.2020.1808896.

[34] M. Dhakal, A. Chhetri, A. K. Gupta, P. Lamichhane, S. Pandey, and S. Shakya, "Automatic speech recognition for the Nepali language using CNN, bidirectional LSTM and ResNet," in *5th International Conference on Inventive Computation Technologies, ICICT 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 515–521. doi: 10.1109/ICICT54344.2022.9850832.

[35] Institute of Electrical and Electronics Engineers, IEEE Computer Society, IEEE Circuits and Systems Society, IEEE Communications Society, and IEEE Signal Processing Society, *2020 IEEE International Conference on Multimedia and Expo (ICME) : 06-10 July 2020, London, UK*.

[36] Z. T. Liu, A. Rehman, M. Wu, W. H. Cao, and M. Hao, "Speech Personality Recognition Based on Annotation

Classification Using Log-Likelihood Distance and Extraction of Essential Audio Features," *IEEE Trans Multimedia*, vol. 23, pp. 3414–3426, 2021, doi: 10.1109/TMM.2020.3025108.

[37] S. Song, S. Jaiswal, E. Sanchez, G. Tzimiropoulos, L. Shen, and M. Valstar, "Self-Supervised Learning of Person-Specific Facial Dynamics for Automatic Personality Recognition," *IEEE Trans Affect Comput*, vol. 14, no. 1, pp. 178–195, Jan. 2023, doi: 10.1109/TAFFC.2021.3064601.