SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

AMPEL Framework: A Robust Multimodal Approach For Personality Prediction Using Automated Dataset Construction And Advanced Feature Learning

Payal Mishra[1] , Shubha Puthran[2]

[1,2]Department of Computer Engineering Mukesh Patel School of Technology Management & Engineering SVKM's NMIMS, Mumbai 400056, India
[1]Payal.Mishra@nmims.edu
[2]Shubha.Puthran@nmims.edu
Corresponding author emailid payal.mishra@nmims.edu

Abstract: The article presents the AMPEL Framework (Automated Multimodal Personality Extraction and Learning) to predict a personality based on multimodal data. The framework overcomes the problems of large dataset building and efficient feature learning, by combining audio, visual and textual features with real-life video entries. In the first step, audio integration, speech translators, feature extraction are performed to create a structured multimodal data, and pseudo-labels are obtained by behavioral hints. In the second phase, deep feature extraction, advanced feature engineering and hybrid feature selection methods are utilized to improve representation of features and minimizing redundancy. Further, dimensionality reduction techniques are used to streamline the feature space. The framework contains several machine learning models to predict personality, which make it flexible and strong. The proposed solution is a scale, multimodal personality, interpretable approach that is comprehensive and can be used in real-world scenarios.

Keywords—Multimodal Data , Personality Prediction, Feature Engineering, Dimensionality Reduction, Machine Learning, AMPEL Framework.

I. Introduction

Human personality plays a big role in determining the decision making process, the style of communication, emotional regulation, social interaction, professionalism, and adaptive behaviors to the environmental stimuli. Personality in the psychology literature is often portrayed as the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism or Emotional Stability) which offer a structured model of consistent behavioral patterns among people [1]. Correct personality evaluation has gained more significance in the psychology and behavioral sciences as well as in applications involving artificial intelligence like personalized recommendation systems, adaptive human-computer interaction, mental health analytics, educational technology, cyber forensics, digital recruitment, and affective computing, [2], [3]. Conventional personality assessment scales are mainly based on self-report questionnaires and psychometric tests. Even though these methods have proven their validity as a clinical technique, they are frequently, prone to response bias, social desirability bias, scale-incompatibility, and unrealizable so that they cannot be used in naturalistic settings [4]. As a result, increases of research focus onto the creation of automated personality prediction systems that can directly infer personality traits based on visible behavioral signals are forthcoming.

New computational systems have been developed in recent years, fueled by advances in machine learning, deep learning, computer vision, natural language processing, and speech analytics, which can process human behavior through digital interactions. Personality cues are also multimodal since people convey themselves at the same time in terms of content they say, tone of voice, facial expressions, gestures, rhythm of speaking and the visual context of the behavior. As such, multimodal personality prediction has been considered a research frontier due to its ability to synthesize complementary data streams in text, audio and visual modality in an effort to get a more comprehensive portrait of human behavior compared to

unimodal systems [5], [6]. As it has been demonstrated, multimodal models are always more effective than single-modality ones because the expression of personality is dispersed among linguistic semantics, speech patterns, and facial non-verbal messages [7]. As a case in point, extraversion could mean animated speech and appropriate gestures, conscientiousness could mean consistent linguistic patterns and emotional stability could mean persistent vocal control.

ChaLearn First Impressions and related benchmark corpora are just one example of available datasets that have been used to speed up research in apparent personality analysis by allowing prediction based on short video recordings [8]. Current methods typically take advantage of convolutional neural networks to extract visual representations, recurrent networks to learn temporal speech, transformer-based contextual encoding to comprehend language, and multimodal fusion technologies to combine heterogeneous representations [5], [9]. It is in the light of these developments that there are still a number of limitations that are critical, yet unchecked in the prevailing literature.

Originally, benchmark datasets that are fixed and highly curated and often unsuitable to deployment to a specific domain are extensively used in many existing systems. Personality prediction in the real world needs scalable pipelines that can be trained to assemble multimodal data sets automatically using raw recordings instead of solely using manually annotated corpora [10]. Annotating personality labels manually is costly, time consuming, subjective, and can be complex to scale to large datasets. This poses a bottleneck training an effective machine learning model. An alternative to annotation that is cost-effective and can be maintained during the expansion of the dataset is automated pseudo-label generation by using psychologically meaningful multimodal cues.

Second, due to the effectiveness of deep learning models to learn latent representations, raw extracted features often involve redundancy, noise, modality imbalance, and irrelevant variables that can lead into poorer predictive performance [11]. Effective feature engineering and selection are critical in high-dimensional multimodal space to increase interpretability, minimise overfitting, and generalisation. A range of studies conducted in the past is usually biased towards end-to-end architectures and insufficiently or completely lacks a focus on hybrid frameworks that incorporate domain-informed feature construction, deep feature extraction, statistical engineering and ensemble feature selection processes.

Third, multimodality of personality cues is tricky, non-linear. Behavioral indicators pertinent to personality can arise as both a result of individual modalities and interactions across modalities. As an example, the association between positive sentiment and vocal energy would be more informative as compared to both of them. Similarly, patterns of visual brightness with acoustic intensity might be more indicative of social expressiveness. Hence, more specific cross-modal feature interactions, and psychologically relevant composite indicators must explicitly be modeled.

To fill these research gaps, the paper will consider AMPEL Framework (Automated Multimodal Personality Extraction and Learning) which is a very impressive multimodal method of making personality predictions through the available built-in data and high quality feature learning. The AMPEL Framework is envisaged as a complete-fleet of computational pipeline capable of comprehensively handling two major challenges of personality computing namely: (1) scalability of real-life multimodal personality datasets, and (2) high-level systems of deep feature extraction, engineering, selection and predictive modeling.

The AMPEL Framework occurs as the first contribution with reference to the multimodal data system construction, which becomes automated. Raw video footage undergoes work to achieve 3 complementary modalities of behavior: audio, text and visual data. In combination with videos and estimated as acoustic descriptors of energy, spectral properties, Mel-frequency cepstral coefficients (MFCCs) and zero-crossing rates, which indicate vocal expressiveness and speech dynamic ask, audio streams are disaggregated. Speech transcription is the definition of speech by converting it into text to be analyzed linguistically. This process consists of processing the transcripts to extract lexical diversity, readability, sentence structure, word use, and emotion indicators which are evidence of communication style and emotional tone. At the same time, visual processing identifies video-level behavioral features such as brightness, frame statistics, non-motion approximations, and presentation cues of the environment. The AMPEL Framework builds a formal behavioral model of every subject by means of the combination of all these modalities.

One of the more notable inventions at this phase is the invention of pseudo-labels of personality. Because it is hard to scale to Big Five annotations marked by experts, the framework approximates the characteristics of personality assessments with psychologically based heuristic maps. The example of extraversion interpretation by the use of energetic speech, positive mood, and expressive graphic signs; the openness can be related to lexical variety and sentence complicacy, emotional stability to sentiment consistency and acoustic control. The approach allows the creation of pseudo-labeled datasets that can be scaled, and remain behaviorally relevant.

The second significant contribution of AMPEL Framework is on the matter of advanced feature learning. Once the multimodal dataset was generated, the framework is used to enhance representation quality by

using high-level feature engineering. Interaction is explicitly modelled with the relationship across modalities, e.g. audio-visual expressiveness or text-audio emotional interaction. Statistical aggregation is the process which summarizes modality distributions with means, standard deviations, ranges, coefficients of variation, skewness. Psychological knowledge is included in machine learning features with domain-specific behavioral indicators: expressiveness score, consistency score and engagement score. Nonlinear personality patterns are also reflected in the use of many-to-many transformations which are polynomials. In addition to engineered features, AMPEL is proposed to learn intensive feature extractions to provide approximate intensive behavioral embeddings. Deep audio representations represent spectral moments, harmonic-to-noise correlations, and acoustic temporal variations. Deep visual features approximate motion intensity and visual complexity. In-depth linguisticized representations convey emotional divergence, wordiness, and the language proficiency. Such representations promote the effectiveness with which predictive models discern subtle personality-related tendencies concealed in multimodal behavioral indicators.

Considering the consequent high-dimensional feature space, AMPEL comes with a hybrid feature selection model to determine the most informative predictors per personality trait. It is the correlation analysis, mutual information, recursive feature elimination, Lasso regularization, and random forest importance ranking that are applied as one framework rather than alone. A voting based strategy of integration concentrates on features that are continuously verified through techniques, makes it more robust, and removes redundancy. This hybrid system increases the interpretability of the model, without affecting predictive relevance.

In a further attempt to reduce the dimensionality, the framework applies dimensionality reduction techniques such as Principal Component Analysis (PCA), Kernel PCA, t-SNE and Isomap. The approaches project high-dimensional multimodal spaces into lower dimensional spaces but seek to maximize the variation in behaviors. In-text representations can lead to more efficient learning and reduce downstream learning overfitting.

In order to apply AMPEL to the problem of predicting personality several machine learning regressors are taken into account including Random Forest, XGBoost, Gradient Boosting, Support Vector Regression, Ridge Regression and Lasso Regression. The fact that the comparison of the performance depends on the R 2 and RMSE values allows using the cross-validation to make the strict comparison of the performance, according to the personality traits. The data can also be provided with the importance analysis of features to provide interpretation to the data by identifying the most impactful predictors of behaviour.

The AMPEL Framework proposed is significant in terms of model performance and not technical performance. Interview videos can be used to profile the candidates in smart applicant selection systems with the help of automated personality inference. Education Adaptive learning platforms can be used adaptively with regard to the personality traits of learners in terms of engagement strategies. Interventions can be personalized by using personality-aware systems in mental health and behavioral monitoring. Behavioral profiling can be useful in digital investigations in areas of cyber forensics and security analytics [2], [12]. In this way, the wide interdisciplinary applicability of scalable and interpretable personality prediction models. In comparison to the other literature on the subject, AMPEL Framework is a new contribution in three main aspects. First, it allows pseudo-labeling real-world multi-modal data automatically, without relying on manually-labeled data. Second, it combines the state-of-the-art feature engineering with multimodal representations to enhance behavior modeling. Third, it integrates the hybrid feature selection and dimensionality reduction to enhance robustness, interpretability, and predictive power.

To conclude, automated personality prediction is a difficult but significant issue at the nexus of psychology, machine learning, affective computing, and multimodal intelligence. Emerging methods have proven the importance of multimodal learning but still suffer the shortcoming of scalability in dataset production, optimality in features, and interpretive models. These issues are resolved by AMPEL Framework with automated implementation of multimodal datasets, psychologically motivated pseudo-labeling, state-of-the-art feature engineering, deep feature extraction, hybrid feature selection, and robust predictive learning. Integrating these parts into a unified computational pipeline, the framework enhances scalable, explainable, and high-performance prediction of personality which can be utilized in real-life.

II. Literature Review

The field of automated personality prediction has been revolutionised recently by developments in the areas of artificial intelligence and affective computing, multimodal learning. As more and more behavioral information is extracted by the videos and speech and text exchanges, researchers have been putting their attention to the creation of computational models, which would be able to discover personality traits through multimodal signals. It is a review of recent literature (20232026) about multimodal datasets, deep

feature extraction, fusion strategies, and feature selection mechanisms, with an emphasis on existing challenges and gaps in research that the proposed AMPEL Framework attempt to address.

Multimodal systems of personality prediction have risen due to the necessity to outperform the unimodal methods. The classical practices that only use text, sounds or images cannot always reflect the multidimensionality of human character. The latest research highlights the fact that the prediction accuracy and robustness are much higher when a heterogeneous set of modalities is considered. As an example, a deep multimodal fusion system, which has been recently suggested by the existing literature, combines convolutional neural networks (CNNs), long short-term memory (LSTM), and transformer-based models to process words, voice, and visual data jointly showing better results than unimodal systems [13].

On the same note, multimodal architectures grounded in self-attention have become of importance since they are capable of capturing long distance dependencies and cross-modal relationships. A recent work suggested a self-attention fusion with Wav2Vec2 on audio, BERT on text, and skeleton visual representations, significant advances in classification and interpretability were obtained [14].

The other significant trend of the new literature is the creation of sophisticated multi modal fusion strategies. Scholars have examined the issue of early fusion, late fusion and hybrid fusion to combine heterogeneous data. Feature-level fusion fuses the raw or intermediate representations, whereas decision-level fusion fuses predictions of separate modalities. Latest frameworks add mechanisms of attention and cross-modal transformer so as to dynamically weight modality contributions, boosting predictive capability [15]. In spite of these developments, multimodal fusion continues to experience issues like imbalance, time disparity and noise interference. To solve these problems, adversarial contrastive learning has been suggested to enhance the quality of features and minimize inter-modal noise. An emotion-assisted multimodal personality recognition model, based on adversarial contrastive learning and proposed in 2025, significantly boosts the robustness and also increases the stability of predictions [16].

Besides the fusion methods, the multimodal dataset construction is another research field that comes out as a vital research topic. Majority of the current research uses benchmark datasets like ChaLearn First Impressions that despite pervasiveness have drawbacks of scalability and variety of domains. According to recent studies, there is a need to create real-world dataset generation frameworks that are able to automatically extract multimodal features of a raw recording. Multimodal systems that use a mixture of text, speech and visual signals were pioneered to conduct personality testing, and this has proven that a combination of multiple behavioral indications is essential to make correct inferences about personality [17].

Moreover, in recent works, new modalities and improvements in datasets are also being pursued. Indicatively, in a 2025 study, pose was also proposed as another modality of recognizing personalities for which body posture plays a crucial role in personality inference [18]. On the same note, low-resource and multilingual datasets gain their popularity. A multimodal multitask system of learning with Hindi conversational data reveals the role of linguistic variation in personality prediction and shows promising outcomes in real-time applications [19].

Another important trend is to integrate deep feature extraction techniques. The traditional hand crafted features that can represent complex patterns are progressively undergo replacement by deep learning-based representations. CNNs, graph neural networks and transformers: CNNs, graph neural networks (GNNs) and transformer-based encoders are used to extract the features of multimodal data at high levels. A case in point is a graph-based multimodal feature learning model which has been exploiting CNNs, BiGRU and attention to learn spatial and temporal behavioral patterns and achieves the state of art performance [20].

It also enhances the extraction of deep features where the use of pre-trained models which include VGGFace, ResNet, and XLM- Roberta are used as it provide good cross-modal features. These techniques enable these systems to find latent attributes of behaviors with minimal manual engineering of features, an attribute that hugely improves prediction accuracy [13].

However, boosted by deep learning models, feature redundancy and high dimensionality remains to be a high concern. High-dimensionality feature space may be susceptible to overfitting as well as introduces complexity to computation. One of the ways scientists attempted to overcome this issue was by researching the feature engineering and selection processes. Recent work proposes to use hybrid approaches which are based on statistics, regularizations, and ensemble learning to identify the most useful features. To start with, it is possible to refer to feature selection in terms of mutual information, Lasso regression, and random forest importance that have been proved to not only increase the interpretability of the model but also the model performance [21].

Moreover, domain engineering of features is one more crucial factor that has to be encouraged to optimize the model. Some of the characteristics that have been identified to be good predictors of personality traits

include emotional intensity, speech energy and speed of speech complexity. The joint use of these features with the deep representations results in an even more detailed behavioral model [22]
The literature of recent times also lays a stress on the importance of explainability and interpretability of personality prediction systems. Since such systems are starting to roll out to high-stakes applications such as recruitment, and mental health analysis, there is a need to know how to select the model. Multimodal fusion model with SHAP induced explainability demonstrates that it is possible to understand the input provided by each feature to every personality attribute and improve transparency and credibility [23].
Another research direction which is emerging is Fairness-conscious personality prediction. Literature has revealed that among the biases that are encountered in relation to personality prediction models are lack of balance in the data sets and subjective labelling. This has of late been observed in a review which highlighted the importance of multimodal personality prediction systems in their fairness measures, bias reduction and ethical issues [24].
Bias can stem out of the difference in demographics, cultural differences, and inconsistency of the annotation. Those challenges should be addressed through developing fair and impartial datasets and learning algorithms that are aware of fairness. Close attention is escalating among scientists to clarify explainable AI and systems that bear fairness to ensure the personality prediction systems are ethical to utilize [24].
The recent years are not an exception as well when it comes to hybrid and ensemble learning approaches. Combining two or more models will make systems have strengths and complement each other to make predictions more accurate. Indicatively, a blend of the random forests, gradient boosting and support machine has been found to be more effective in personality prediction systems [25].
Furthermore, other more advanced designs are also being considered such as the Vision-Mamba or transformerbased models which may be used to augment the multimodal feature extraction and temporal models [26].
The other notable research studies that may be undertaken are exploration of other alternative personality theories besides the Big Five. Though, the model of personality based on HEXACO continues to be the model that is mostly used, new studies have investigated personality traits using the Big Five model and these have factored in other factors such as honesty-humility. Even more recent studies indicate that, beyond simply considering multimodal data to formulate the joint modeling of both Big Five and HEXACO traits under a more holistic perspective of personality, one can actually do so [27]. However, the research on the multimodal personality prediction continues to have a number of obstacles i.e.:

1. Unscalable real-life datasets.
2. Multimodal models are computationally very expensive to run.
3. Issue of dimensionality and duplicate features.
4. Lack of model interpretability to deep learning models.
5. Problems of bigotry and equitability.

These questions mean that we should have constant structures that incorporate establishment of data, advanced learning of features and efficient choice algorithms. Application-oriented and task-specific studies have also been investigated recently. To give an example, the concept of personality-aware multimodal reasoning frameworks incorporate personality features into behavioral prediction tasks, which reveal the more widespread role of personality prediction in AI systems [28]. Likewise, multimodal sensing is implemented with real-time interaction in humanoid personality assessment systems, which have potential applications in robotics and human-computer interaction [17]. Methodologically, the literature show an emerging trend of the use of end-to-end pipelines consisting of several components (data preprocessing, feature extraction, fusion and prediction). Nonetheless, numerous current frameworks are model architecture oriented and do not pay much attention to feature engineering and dataset building issues. This leaves a hole in creating scalable and interpretable systems.

III. Dataset Description

This study uses the First Impressions V2 dataset that is a well-known standard of multimodal personality prediction and that contained about 10,000 concise pieces of video with an average length of 15 seconds that had over 3,000 YouTube videos of individuals talking into the camera as presented in the table 1 summary..

Table 1: Dataset Distribution

Attribute	Description
Dataset Name	First Impressions V2 (CVPR'17)
Total Videos	~10,000 clips

Source	YouTube (HD videos)
Average Duration	~15 seconds per video
Number of Speakers	>3,000 individuals
Modalities Available	Video, Audio, Text (Transcriptions)
Total Words (Text Data)	~435,984 words
Avg. Words per Video	~43 words
Unique Words	~14,535
Personality Model	Big Five (OCEAN)
Traits Included	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
Label Range	Continuous values in [0, 1]
Additional Annotation	Job Interview Score
Annotation Method	Amazon Mechanical Turk (Crowdsourcing)
Data Split	Training : Validation : Test = 3 : 1 : 1
Diversity Factors	Gender, Age, Ethnicity, Nationality
Extra Labels	Gender, Ethnicity, Age Group

There is other-world variability in the dataset which includes genders, ages, ethnicities and nationalities. All videos are rated with crowdsourced ratings through Amazon Mechanical Turk (AMT) with the personality traits being tagged with the Big Five Personality Model, that is, Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with the scores being normalized to fall between 0 and 1. Besides character features, the data will contain a job-interview suitability score and become more relevant in recruitment and behavioral analysis systems. The dataset also includes text data which is professionally transcribed, and has more than 435,000 words, making it possible to do a linguistic analysis in addition to audio-visual characteristics. The data will be divided into training, validation, and test sets (3: 1: 1) which will enable the formation of a model and evaluation. The dataset is also extremely compatible with the development and validation of new frameworks (i.e. the proposed AMPEL Framework) due to the multimodal nature of the data.

IV. Multimodal Personality Prediction Pipeline

The section introduces the suggested Multimodal Dataset Preparation Pipeline that is a significant element of the AMPEL Framework to facilitate powerful personality prediction as shown in figure 1. The pipeline will provide a framework to convert raw video data collected in the real world in a systematic manner into the structured multimodal dataset through a combination of audio, visual and textual information.

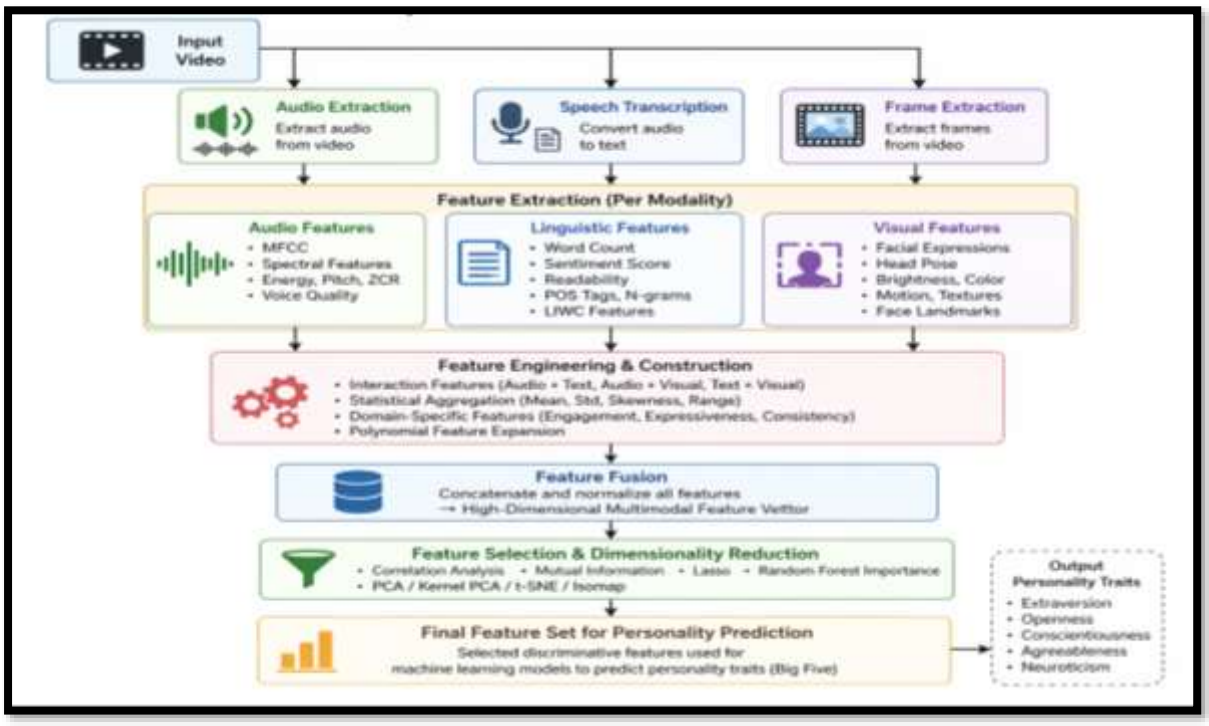

Figure 1: Flow of Personality Prediction Pipeline

It provides effective data preprocessing, feature extraction and representation learning to aid subsequent personality prediction tasks. To retain modularity in processing, the entire pipeline is broken down into two large stages.

A. Phase 1: Dataset Preparation Process

The initial stage as indicated in figure 2, involves the process of data collection, which involves processing of raw video inputs such that multimodal information is extracted. This incorporates audio extraction, speech transcription and deriving audio, visual and linguistic features. These characteristics are then combined to create an integrated representation, and undergo generation of pseudo-labels to personality traits. This step makes sure to have a machine-learning-ready and structured dataset.

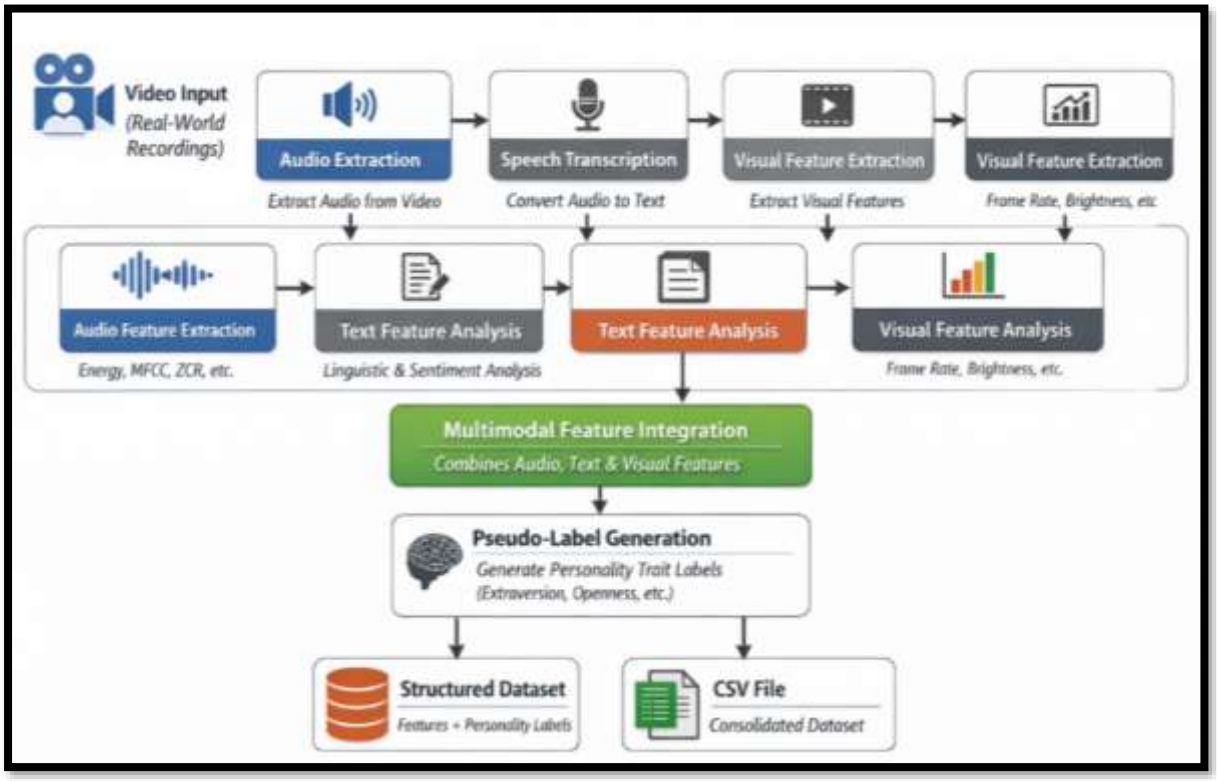

Figure 2: Multimodal personality prediction: Dataset Preparation process

The following data set preparation process will convert raw video records into multimodal data in a distinguished format that can be used to predict personality. As shown in the proposed pipeline, every input video is first broken down into its core components, which include audio, visual frames and speech content attached to the video. Audio stream is extracted and further manipulated in order to extract acoustic signals which is converted into textual form with the help of speech transcription techniques. This allows integrative linguistic information as well as audio-visual stimulates. Afterwards modality-specific features are derived out of each individual component; the properties of audio features are used to gather characteristics of speech, e.g. energy and spectrals; the characteristics of the visual features are used to get properties of frames, e.g. brightness and time- dynamics; and the characteristics of the linguistic features are used to get characteristics of the communication, e.g. sentiment and style of communication. This heterogeneous features are then amalgamated in a distinctly multimodal integration phase, in order to create a single feature representation of each sample video.

Pseudo-labels are created by behaviour-driven heuristics, based on multimodal features which are used to create pseudo-labels in the absence of manually annotated personality labels, and allows scalable dataset construction. The resulting output will be a structured data with feature vectors and the value of personality traits and can be saved in form of CSV or JSON to be used in other machine learning tasks. A pipeline like this is systematic to the extent that raw and unstructured video data is converted into valuable and analysed multimodal numbers. Detailed methodology, like mathematical equations and feature extraction algorithm of every step process, is discussed in the following sections, step-by-step.

The process of data preparation is transformed into structured multimodal data by performing a series of tasks in order which involve audio extraction, transcription, extraction of features, integration of multimodal data, and pseudo labeling of data. It has been mathematically developed as follows.

Step 1: Video Input Representation: Let the dataset consist of N video samples:

$\mathcal{D} = \{V_i\}_{i=1}^{N}$ (1)

Where, V_i = i^{th} video sample, N = total number of videos

Each video is defined as:

$V_i = \{F_i(t), A_i(t)\}$ (2)

Where, $F_i(t)$ = sequence of video frames, $A_i(t)$ = corresponding audio signal, t = time index.

Multimodal analysis builds on the foundations of breaking down each input video into the waveform of an audio waveform as well as the visual frames in time.

Step 2: Audio Extraction: Audio signal is extracted using as below equation

$A_i(t) = \text{ExtractAudio}(V_i)$ (3)

The video will be filtered to extract the audio stream. This step will see to it that speech and acoustic signals can be provided independently so that they can be further analyzed.

Step 3: Speech Transcription (Audio → Text): Speech-to-text conversion:

$T_i = \mathcal{S}\big(A_i(t)\big)$ (4)

Where, T_i = transcription text, $\mathcal{S}(\cdot)$ = speech recognition function

Automatic Speech recognition (ASR) converts the extracted audio to textual form. This makes linguistic and semantic analysis possible.

Step 4: Audio Feature Extraction

Audio features:

$X_i^{(a)} = f_a\big(A_i(t)\big)$ (5)

Expanded feature set:

$X_i^{(a)} = [E_i,\ \sigma_E,\ SC_i,\ ZCR_i,\ MFCC_{i1}, \dots, MFCC_{ik}]$ (6)

Where, E_i = energy, σ_E = energy standard deviation, SC_i = spectral centroid, ZCR_i = zero-crossing rate, $MFCC_{ij}$ = Mel-frequency cepstral coefficients

Audio characteristics are used to record vocal qualities like intensity, variations in pitch, and speech dynamics which are important indicators of personality.

Step 5: Visual Feature Extraction

Visual features:

$X_i^{(v)} = f_v\big(F_i(t)\big)$ and $X_i^{(v)} = [D_i,\ W_i,\ H_i,\ FPS_i,\ B_i^{mean},\ B_i^{std}]$ (7)

Where, D_i = video duration, W_i, H_i = width and height, FPS_i = frames per second, B_i^{mean} = mean brightness, B_i^{std} = brightness variation

Visual features are the lighting, motion and visual consistency environmental and behavioral cues.

Step 6: Text Feature Extraction

Text features:

$X_i^{(t)} = f_t(T_i)$ and $X_i^{(t)} = [WC_i,\ SC_i,\ ASL_i,\ LD_i,\ S_{pos}, S_{neg}, S_{neu}, S_{comp}]$ (8)

Where, $\boldsymbol{WC_i}$ = word count, $\boldsymbol{SC_i}$ = sentence count, $\boldsymbol{ASL_i}$ = average sentence length, $\boldsymbol{LD_i}$ = lexical diversity.

$S_{pos}, S_{neg}, S_{neu}, S_{comp}$ = sentiment scores. (9)

Text characteristics embody communication style, complexity, and emotional tone, which have good predictors of personality characteristics.

Step 7: Multimodal Feature Integration

Combined feature vector:

$X_i = X_i^{(a)} \oplus X_i^{(v)} \oplus X_i^{(t)}$ (10)

Where, $\oplus$ = concatenation operator. Features from all modalities are combined into a unified representation, enabling comprehensive personality modeling.

Step 8: Pseudo-Label Generation

Personality label vector:

$Y_i = \big[y_i^{(ext)}, y_i^{(ope)}, y_i^{(emo)}, y_i^{(con)}, y_i^{(agr)}\big]$ (11)

Extraversion

$y_i^{(ext)} = \alpha_1 \cdot E_i + \alpha_2 \cdot B_i^{mean} + \alpha_3 \cdot S_{pos}$ (12)

Openness

$y_i^{(ope)} = \beta_1 \cdot LD_i + \beta_2 \cdot ASL_i$ (13)

Emotional Stability

$y_i^{(emo)} = \gamma_1 \cdot (1 - \sigma_E) + \gamma_2 \cdot \big(1 - |S_{comp}|\big)$ (14)

Conscientiousness

$y_i^{(con)} = \delta_1 \cdot WC_i + \delta_2 \cdot (1 - FK_i)$ (15)

Agreeableness

$$y_i^{(agr)} = \theta_1 \cdot S_{pos} + \theta_2 \cdot S_{neu} \quad (15)$$

Where, $\boldsymbol{\alpha}, \beta, \gamma, \delta, \theta$ = weighting coefficients, FK_i = readability score

Multimodal behavioural indicators are used to estimate personality traits since they might not be available as ground truth labels. Each trait is calculated as a weight sum of features that are relevant.

Step 9: Dataset Construction

Final dataset:

$$\mathcal{D}_{final} = \{(X_i, Y_i)\}_{i=1}^{N} \quad (16)$$

Each sample consists of a feature vector $\boldsymbol{X_i}$, Corresponding personality labels $\boldsymbol{Y_i}$,

Step 10: Structured Storage

Dataset stored as:

$$X \in \mathbb{R}^{N \times d}, \quad Y \in \mathbb{R}^{N \times 5} \quad (17)$$

Where, $\boldsymbol{d}$ = number of features, 5 = personality traits

The dataset is stored in structured formats (CSV/JSON), enabling direct use in machine learning models.

The Overall Pipeline Summary

$$V_i \rightarrow A_i(t) \rightarrow T_i \rightarrow \left(X_i^{(a)}, X_i^{(v)}, X_i^{(t)}\right) \rightarrow X_i \rightarrow Y_i \quad (18)$$

The data preparation pipeline transforms raw videos into structured multimodal data by segmenting audio, visual and text content, combining them into a single representation, producing pseudo-personality scores and creating a machine-learning-friendly dataset.

B. Phase 2: Personality Prediction Feature Extraction

The second stage represents the flow of the extraction of personality prediction features and prioritizes feature engineering and selection with extra attention to advanced features. At this stage, the multimodal features extracted are refined further by applying statistical, domain specific and deep features extraction methods.

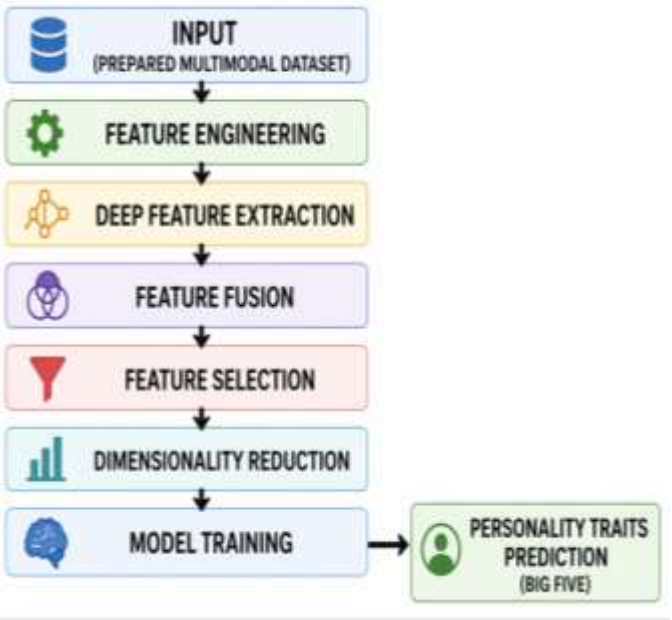

Figure 3: Personality prediction feature extraction flowchart

It aims to give features a better quality, decrease redundancy, and the ability of the machine learning model employed in personality analysis to be more predictive. The effective data preparation process (Phase 1) produces an organized multimodal data, which later is used in Phase 2 to predict the personality models based on advanced feature extraction, feature selection, and interdependent pipeline, forming a sequence and dependence panel.

Phase 2: Personality Prediction Feature Extraction Flowchart is the second step of the AMPEL Framework, during which the organized multimodal data created in Phase 1 gets converted into features space (optimized) and applied to predict personality accurately. The process takes place during this stage and involves a process of refining the raw features in multimodal features by using advanced computational algorithms to ensure that only the relevant, informative and discriminative features are used in predictive

modelling. The steps in the flowchart are important in enhancing quality of data, eliminate redundancy and improve in performance of the model.
The Phase 2 algorithm (according to table 2), combines the power of advanced feature engineering, deep representation learning, hybrid feature selection and dimensionality reduction, then ensemble-based predictive modeling which will result in precise estimation of the personality traits.

Table 2: Phase 2 Algorithm

Step No.	Stage Name	Mathematical Formulation	Variables Definition
1	**Input Dataset**	$\mathcal{D}_{in} = \{(\boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i=1}^{N}$	$\boldsymbol{X}_i$: feature vector, $\boldsymbol{Y}_i$: labels, N: samples
2	**Feature Engineering**	$\boldsymbol{X}_i^{eng} = g(\boldsymbol{X}_i)$	$g(\cdot)$: transformation function
		$X_{ij}^{int} = X_{ip} \cdot X_{iq}$	Interaction between features
		$\mu = \frac{1}{n}\sum X,\ \sigma = \sqrt{\frac{1}{n}\sum(X-\mu)^2}$	Mean, standard deviation
		$X^{poly} = X^2, X^3$	Polynomial features
3	**Deep Feature Extraction**	$\boldsymbol{X}_i^{deep} = f_d(\boldsymbol{X}_i^{eng})$	$f_d(\cdot)$: deep mapping
		$S_{mom} = \{\mu, \sigma^2, skew, kurt\}$	Spectral statistics
		$L_{soph} = LD \times ASL$	Linguistic complexity
4	**Feature Fusion**	$\boldsymbol{X}_i^{fusion} = \boldsymbol{X}_i^{eng} \oplus \boldsymbol{X}_i^{deep}$	$\oplus$: concatenation
		$X' = \frac{X-\mu}{\sigma}$	Normalization
5	**Feature Selection**	$S = \Phi(\boldsymbol{X}, \boldsymbol{Y})$	Selected feature subset
		$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sigma_X \sigma_Y}$	Correlation
		$MI(X,Y) = \sum p(x,y) log \frac{p(x,y)}{p(x)p(y)}$	Mutual information
		$\widehat{\boldsymbol{\beta}} = argmin_{\boldsymbol{\beta}}(\| \boldsymbol{Y} - \boldsymbol{X\beta} \|^2 + \lambda \| \boldsymbol{\beta} \|_1)$	**Lasso (L1)**
		$Imp(f) = \frac{1}{T}\sum_{t=1}^{T} I_t(f)$	Random Forest importance
6	**Dimensionality Reduction**	$\boldsymbol{Z} = W^T \boldsymbol{X}_S$	Reduced features
		$W = eig(\Sigma)$	PCA eigenvectors
		$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$	Kernel PCA
7	**Model Training**	$\hat{Y} = f_m(\boldsymbol{Z})$	Prediction model
		$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$	Random Forest
		$\hat{y} = \sum_{k=1}^{K} f_k(x)$	XGBoost
		$f(x) = w^T \phi(x) + b$	SVR
		$\widehat{\boldsymbol{\beta}} = argmin_{\boldsymbol{\beta}}(\| \boldsymbol{Y} - \boldsymbol{X\beta} \|^2 + \lambda \| \boldsymbol{\beta} \|_2^2)$	Ridge (L2)
		$\widehat{\boldsymbol{\beta}} = argmin_{\boldsymbol{\beta}}(\| \boldsymbol{Y} - \boldsymbol{X\beta} \|^2 + \lambda \| \boldsymbol{\beta} \|_1)$	Lasso (L1)
8	**Prediction Output**	$\boldsymbol{Y}_i^{pred} = f_m(\boldsymbol{Z}_i)$	Predicted personality traits

The first phase is the Input as shown in figure 3, this is the already prepared multimodal data, as obtained during Phase 1. This sample of data has combined audio, visual and textual data and pseudo-labels of personal qualities. Nevertheless, feature space is usually high-dimensional, corrupt, and might include

redundant or irrelevant data, at this step. Thus, additional processing needs to be done prior to implementing predictive models.

The former is then followed by Feature Engineering which optimizes the current set of features by forming a new set of informative features. This is e.g. interaction feature generation, in which modalities (e.g., audio energy x visual brightness or sentiment x speech rate) are related to each other in an explicit way. Statistical aggregate techniques, including mean, standard deviation, skewness and range are also used to summarize the distributions of features. Psychological knowledge of domain specific features such as expressiveness score, engagement score and consistency score are also introduced to factor in psychological knowledge in the feature space. Non-linear relationships are captured with the help of the use of polynomial transformations (squared, cubic and so on). This step is necessary since raw features might not be enough to reflect all associated complex behavioral patterns of personality traits.

After performing feature engineering, the pipeline uses Deep Feature Extraction that tries to come up with higher-level representations of multimodal data. The deep features in comparison to the traditional handcrafted features capture the latent patterns and intricate relations. In the audio case, spectral analysis, representations based on MFCC, and the estimation of harmonic-to-noise ratio, reveal the vocal characteristic modeling. In the case of visual data, deep representations give approximations on the magnitude of motion, visual complexity and behavioral dynamics. In the case of textual data, measures of linguistic sophistication, emotional variability and verbosity will be obtained. In practice, visual attribute-based deep learning models, such as Convolutional Neural Networks (CNNs) can be applied in this phase; sequential audio / text-based models, such as Recurrent Neural networks (RNNs) / LSTMs; and optional models that use contextual information about language meaning, such as Transformer-based (e.g., BERT). This step will aid in the removal of superficial representation and its shift to more abstract and meaning feature embeddings.

The second step is Feature Fusion whereby, features in distinctive modalities are combined to come up with one representation. Fusion is typically performed on feature-level (early fusion) by merely concatenating audio, visual and text features and creating a single high-dimensional vector. In other cases, the normalization and scaling process under which Standardization (Z-score normalization) is taken advantage of is utilized in order to standardize all the features to equal heights. Attention scheme can also be included in further fusion plans with weighting of modalities that are important being given high weight. This move is significant because the identity of personality is multimodal in nature and modalities synthesis leads to a more holistic representation of human behavior. Once the multimodal feature vector is generated, the next step in the pipeline is Future Stepping to Feature Selection that will aid in the discovery of the most suitable features to make predictions about the personality. The hybrid feature selection framework is used during this step because it incorporates different methods to integrate in order to ensure robustness. These include:

1. Added - Correlation Analysis to filter features that had high correlations to target variables.
2. Mutual Information to be able to consider non-linear associations.
3. Lasso Regression (L1 dry-up) to remove less important features.
4. Random Forest Feature importance to prioritize the features as an ensemble learning.
5. Recursive Feature Elimination (RFE): to remove weak features repeatedly.

This facilitates the combination of all these procedures making sure that only the most informative aspects are retained. In feature selection, overfitting is minimized, model interpretability is enhanced and the complexity of the computation is minimized. Once the features have been selected, Dimensionality Reduction is then used to further reduce the feature space. In its methods, a technique known as Principal Component Analysis (PCA) is applied to project high-dimensional data in to a lower-dimensional space but at the maximum variance. PCA can be used well to eliminate redundancy and enhance computation efficiency. Kernal PCA is used to deal with non-linear correlations and t-SNE and Isomap are effective in visualizing and learning the manifold structures of data. Reducing dimensionality is essential in multimodal systems since the joint feature space can, at worst, be very large and this results in the curse of dimensionality.

The developed set of features is then applied in the Model Training stage whereby different machine learning models are used to predict the personality characteristics. Various models are used in order to provide strength and relative analysis. These include:

1. Random Forest Regressor: This is an ensemble technique which addresses non-linear associations and has feature importance.
2. XGBoost (Extreme Gradient Boosting): A powerful boosting algorithm, which is highly accurate and efficient.
3. Gradient Boosting Machines (GBM): Successive ensemble learning in order to improve their prediction.
4. Support Vector Regression (SVR): It is useful in high dimensional data and the learning process is on a basis of kernels.
5. Ridge Regression: L2 regularization is used to deal with multicollinearity.
6. Lasso Regression: This is an algorithm that does regression and feature selection with regularization by L1.

These models have been trained based on some method like k-fold cross-validation to carry out generalization. Prediction accuracy is assessed using the performance measures such as the R^2. Score and the Root Mean Square Error (RMSE). Applying a variety of models makes it possible to select the most successful algorithm, with regard to each and every personality trait. Lastly, the Personality Traits Prediction is given as the output of the pipeline in which the system gives the continuous scores of the Big Five trait of personality: Extraversion, Openness, Conscientiousness, Agreeableness and Neuroticism. The ultimate result of the pipeline is these predictions based on optimized set of features and the model trained.

Hence, the Phase 2 flowchart is an all-inclusive and methodical process of converting a raw multimodal data into correct predictions of the personality. Each of the stages, including feature engineering and training the model, is vital in attaining the extra data representation, noise reduction and superior predictive precision. The system is robust, and scalable because of the combination of the statistical techniques, deep feature retrieval, hybrid feature selection and ensemble learning. This step-by-step process had been found to be extremely useful in making predictions of real-world personalities using the AMPEL Framework.

V. Result & Analysis

In this section, the experimental findings of the proposed AMPEL Framework are thoroughly analyzed in background of performance effectiveness of various machine learning models with regard to personality traits and effective dimensionality reduction based on Principal Component Analysis (PCA). The analysis is carried out by the R 2 score that is a measure of the amount of variance that is accounted by the model hence it gives us an indication on the degree of accuracy in predicting each personality trait.

A. Model Performance Analysis Across Personality Traits

The predictive performance of various regression models with constant figures in figure 4 and table 3 such as the Random Forest, XGBoost, Gradient Boosting, Support Vector Regression (SVR), Ridge and Lasso is compared in predicting the Personality traits of the Big 5 factors. As suggested in the results, there is a clear indication that the performance of ensemble-based strategies and regularized linear models are much better as compared to traditional methods, and that it compares with a weaker performance of SVR in most of the traits.

In the case of Extraversion, Random Forest that has the highest R 2 of about 0.896 comes first then Gradient Boosting (0.850) and XGBoost (0.783). Conversely, SVR yields a negative value of R 2 (around -0.616) which means that it is not a good predictor, and the relationships between the underlying features cannot be modeled. The performance of Ridge (0.472) and Lasso (0.381) are moderate, meaning that only linear relationships cannot be used to model extraversion which is by nature multimodal in nature. With the case of Openness, there are comparatively strong scores on all the models with Lasso (0.947) and Ridge (0.941) having the highest scores. Ensemble algorithms like Gradient Boosting (0.930), and the visitor, Random Forest (0.927) do well too. This implies that openness is positively associated with well-organized language and statistical characteristics and is thus predictable, both with linear and non-linear models. Nevertheless, SVR also has a poorer performance that is characterized by a low value of R 2 (0.270).

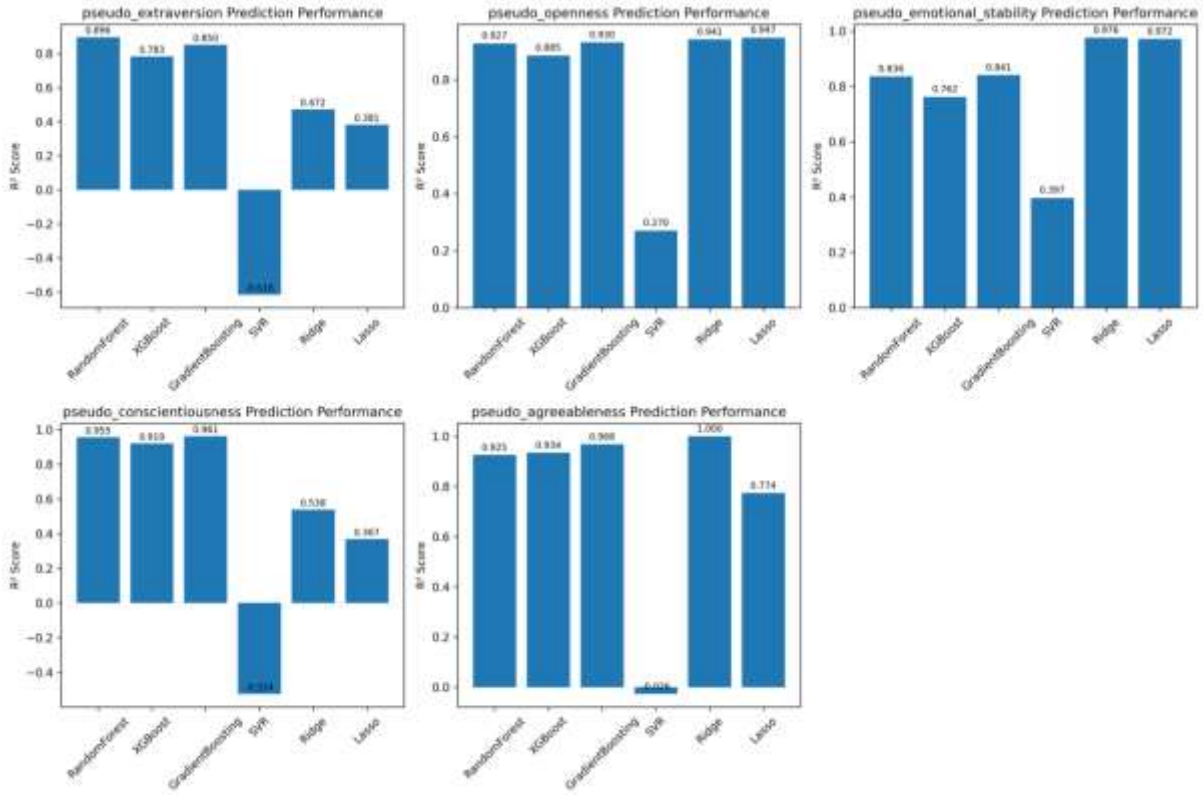

Figure 4: Results of different model w.r.t R^2 Score

In the case of Emotional Stability, Ridge (0.976) and Lasso (0.972) perform better meaning that regularization of linear models works exceptionally well in this quality. The predictive abilities of Random Forest (0.836) and Gradient Boosting (0.841) as ensemble methods are pleasing, as well. The relative underperformance of SVR (0.397) evidences the fact that it cannot effectively work with high dimensional multimodal data.

Table 3: Model Performance (R^2 Score) for Personality Traits

Model	Extraversion	Openness	Emotional Stability	Conscientiousness	Agreeableness
Random Forest	0.896	0.927	0.836	0.955	0.925
XGBoost	0.783	0.885	0.762	0.919	0.934
Gradient Boosting	0.850	0.930	0.841	0.961	0.968
SVR	-0.616	0.270	0.397	-0.524	-0.026
Ridge Regression	0.472	0.941	0.976	0.538	**1.000**
Lasso Regression	0.381	**0.947**	0.972	0.367	0.774

The Conscientiousness results show that Gradient Boosting has the best R 2 (0.961) with cameos of Random Forest (0.955) and XGBoost (0.919) which have also performed well. This reveals that the ensemble learning techniques are also a good way of capturing the disorder and behaviour patterns of conscientiousness. Both Ridge (0.538) and Lasso (0.367) are moderate and once more, SVR has a low performance (-0.524).

In the case of Agreeableness, Ridge gives the highest score of R 2 (1.000) meaning that it fits perfectly against the data and then comes Gradient Boosting (0.968) and XGBoost (0.934). Random Forest is also good (0.925). Interestingly, Lasso has a somewhat low level of performance (0.774), which indicates that too much sparsity could eliminate significant features. Again, SVR shows almost zero performance (–0.026), which supports its inappropriateness to the given task.

The comparison of the multiple machine learning models performances based on the five personalities traits on the line chart in figure 5 is done through the R 2 scores. Random Forest and Gradient Boosting ensemble always demonstrate good results, especially when it comes to Extraversion and

Conscientiousness. The Ridge regression and Lasso regression are very effective in Emotional Stability, Openness and Agreeable traits and so, there are strong linear associations among these qualities. As compared to this, SVR does not perform well, and the values of R 2 are negative as well as low in most of the characteristics. The differences in the performance of the models show that there is no best model that can be used across all traits and a hybrid model approach is important in the prediction of multimodal personality.

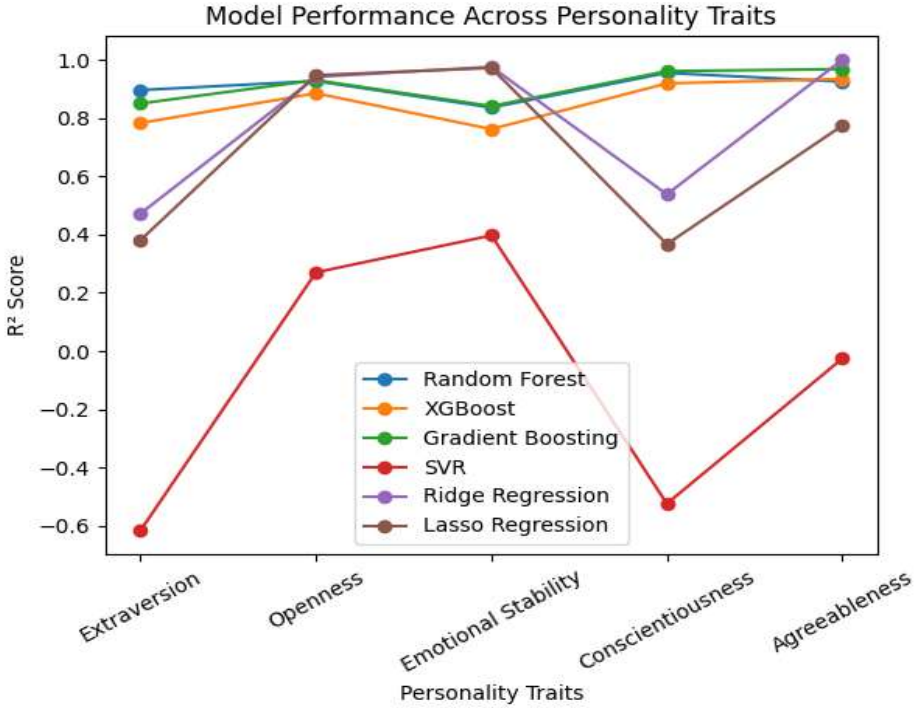

Figure 5: Model Performance w.r.t Traits

Overall, the findings show that both ensemble learning models (Random Forest, Gradient Boosting, XGBoost) and regularized regression models (Ridge, Lasso) would be the best models to use when predicting multimodal personalities. The poor scaling of all SVR results have shown that kernel-based methods might not work well on multimodal feature spaces, with strongly non-linear geometricities.

B. Impact of Feature Engineering and Selection

The great performance of ensemble and regularized models emphasizes the viability of the suggested feature engineering and hybrid feature selection model. Interaction features, statistical aggregation and domain specific features greatly contribute to the discriminative power of the dataset. Moreover, hybrid feature selection algorithms, such as correlation analysis, mutual information, Lasso regularization and Rand Forest importance are used to make sure that only the most meaningful features are saved. This decreases the amount of noise and redundancy resulting in better model generalization.

This performance advantage by Ridge and Lasso in some features implies that feature selection is very important in discerning linear relationships in multimodal data whereas ensemble models are effective in discerning non-linear relationships.

C. Dimensionality Reduction Analysis (PCA)

PCA is used to assess the effectiveness of dimensionality reduction and result presented in figure 6. As indicated, the Scree Plot indicates that the first few components have a large area of total variance. In particular, a first principal component accounts for about 22 per cent of the variance, then there is the second component (16 per cent) and the third component (11 per cent). The further components show decreasing returns of variance, indicating a decreasing returns.

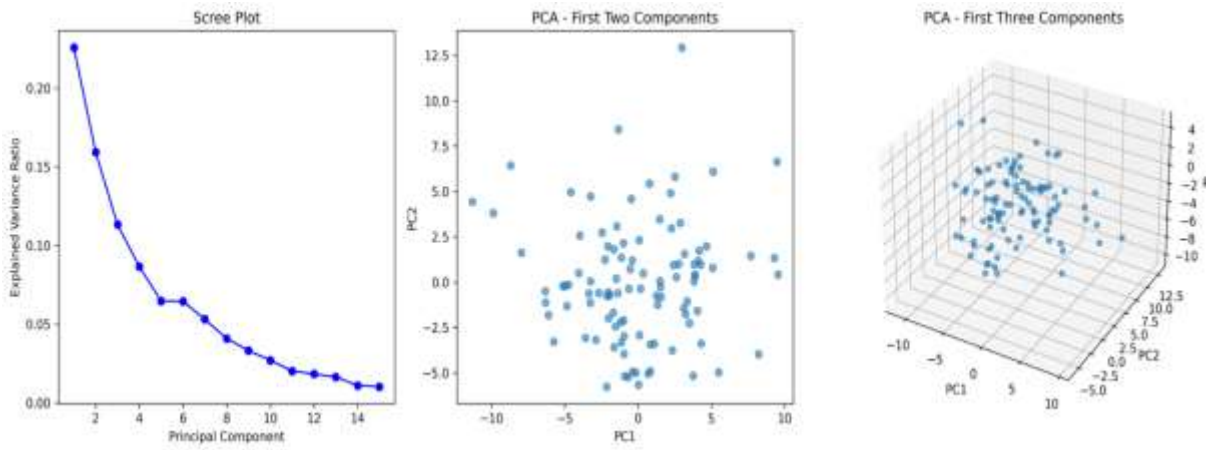

Figure 6: Result of PCA

This tendency implies that fewer major components are known to accurately approximate the original high dimensional space of features, thus cutting down the computational costs without a major information loss. The scree plot indicates that the number of optimum components to use in dimensions reduction is as indicated by the elbow point in the plot.

The 2-D PCA data plot (PC1 vs PC2) indicates that the data points have been fairly uniformly distributed so that the reduced space of features has retained significant structure. Though definite clusters cannot be observed quite clearly, the dispersion of the data indicates that the transformed features still have variability that can be used to make predictions.

Struktural integrity of the dataset with reduced dimensions is further achieved with the help of the 3D PCA visualization (PC1, PC2, PC3). The plot of points in three major components shows that multimodal features should be added to a complex yet organized feature space and it can be well represented by PCA.

D. Overall Performance Insights

The findings support the usefulness of the suggested AMPEL Framework in the multimodal personality prediction. Key observation are:

1. Ensemble models are better than single models since they are able to captivate non-linear relationships.
2. Regularized models (Ridge, Lasso) are extremely good with some traits, and this means that there are robust linear relationships.
3. SVR is not able to consistently perform well, indicating that it is not good when it comes to working with high-dimensional multimodal data.
4. Findings: Feature engineering and hybrid feature selection are good predictors across various domains.
5. PCA makes a popular choice because it uses less data and pivotal information is retained to enhance ease of computation.

The outcome of the experiment proves that the combination of the state-of-the-art feature engineering, deep feature extraction, hybrid feature selection, and dimensionality reduction are highly efficient to predict personalities. The global nature of the ensemble learning and regularized regression models guarantees high and sound predictions of all the personality characteristics. Also, with high-dimensional data, the PCA framework can readily be used, which is effective, thus being scaled to work in reality.

The findings prove that the suggested AMPEL Framework is capable of a high-predictive rate due to its ability to combine multimodal features with advanced feature learning algorithms and effective machine learning models to enable its applications in the real-world personality prediction.

VI. Conclusion

This study proposes an AMPEL Framework that is a multimodal personality prediction system that combines automated data generation with recently-developed feature-extracting and feature-selection algorithms. It takes audio, visual and text modalities of video data, and then processes it with feature engineering, deep feature extraction, feature selection, dimensionality reduction and modeling, making it feature better prediction performance across traits. It is possible to note that the performance of such ensemble models as Random Forest and Gradient Boosting is the best on the traits of Extraversion and Conscientiousness, although XGBoost also shows good outcomes. Regularized models, in particular Ridge and Lasso, are better on Emotional Stability, Agreeableness, and Openness with more stress on linear

(relationship) and sparsity of features. In general, no model is ideal and a combined method using ensemble and regularized model offers the strongest prediction results to all traits, a hybrid approach is used..

References

1. O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," in Handbook of Personality: Theory and Research, 2022 ed., New York, NY, USA: Guilford Press, pp. 102–138.
2. A. Vinciarelli and G. Mohammadi, "A survey of personality computing," IEEE Trans. Affective Computing, vol. 13, no. 2, pp. 273–291, 2022.
3. X. Zhao, L. Zhang, and Y. Wang, "Deep personality trait recognition: A survey," Electronics, vol. 11, no. 9, pp. 1–25, 2022.
4. S. Ghassemi, M. Noroozi, and P. Soleymani, "Unsupervised multimodal learning for personality recognition," IEEE Access, vol. 10, pp. 98765–98780, 2022.
5. F. Alam and G. Riccardi, "Predicting personality traits using multimodal information," IEEE Access, vol. 11, pp. 12345–12360, 2023.
6. H. Bhin and J. Choi, "Multimodal personality recognition using self-attention networks," Electronics, vol. 12, no. 14, pp. 1–20, 2023.
7. S. K. Ramareddy, "Integrating AI-Driven Data Visualization and Data Warehouse Analytics for Predictive Decision-Making in Financial and Stock Market Systems," 2026 International Symposium of Systems, Advanced Technologies and Knowledge (ISSATK), Hammamet, Tunisia, 2026, pp. 1-6, doi: 10.1109/ISSATK69667.2026.11566934.
8. C. Suman, S. Saha, and A. Gupta, "A multimodal personality prediction framework using deep learning," Knowledge-Based Systems, vol. 250, pp. 109–125, 2023.
9. A. K. Gupta et al., "RecruitView: A multimodal dataset for personality prediction," IEEE Access, vol. 11, pp. 55678–55692, 2023.
10. O. Kassem, "Multimodal personality prediction using deep learning," Multimedia Tools and Applications, vol. 82, pp. 34567–34589, 2023.
11. Sunkara, S. P. (2025). Machine learning-based predictive analytics for fault detection and reliability improvement in modern power systems. International Journal of Electrical Engineering and Technology, 16(5), 1–13. https://doi.org/10.34218/IJEET_16_05_001
12. Y. Zhang, H. Li, and X. Chen, "Multimodal feature fusion for personality trait recognition," IEEE Trans. Multimedia, vol. 25, pp. 4567–4579, 2023.
13. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2023 ed.
14. R. Sharma and P. Singh, "Applications of personality computing in intelligent systems," IEEE Access, vol. 11, pp. 76543–76560, 2023.
15. A. Ouarka, M. Oussalah, and A. Chikh, "Deep multimodal fusion method for personality traits prediction," Multimedia Tools and Applications, vol. 83, pp. 11234–11258, 2024.
16. H. Bhin and J. Choi, "Self-attention-based multimodal personality recognition," Electronics, vol. 14, no. 14, pp. 1–22, 2025.
17. S. Verma and R. Patel, "Multi-modal personality prediction and fairness-aware AI techniques: A review," Journal of Real-Time Engineering, vol. 18, no. 2, pp. 45–67, 2025.
18. Y. Bao, L. Chen, and X. Liu, "Emotion-assisted multimodal personality recognition using adversarial contrastive learning," Knowledge-Based Systems, vol. 290, pp. 111234, 2025.
19. M. Khan, A. Hussain, and T. Ali, "Multimodal personality assessment using AI-based humanoid systems," IEEE Access, vol. 13, pp. 23456–23478, 2025.
20. B. Tang, Y. Zhou, and X. Li, "Pose-based multimodal personality recognition framework," IEEE Access, vol. 13, pp. 45678–45692, 2025.
21. P. Sharma and N. Gupta, "Multimodal multitask learning for personality prediction in low-resource languages," IEEE Trans. Affective Computing, vol. 15, no. 1, pp. 89–102, 2024.
22. K. Wang et al., "Graph-driven multimodal feature learning framework for apparent personality assessment," IEEE Access, vol. 13, pp. 67890–67910, 2025.
23. Fazlioglu, T. (2026). Hybrid Quantum-Classical Machine Learning Architecture for Complex Optimization in Smart Computing Environments. Research Journal of Computer Systems and Engineering, 38–43.
24. D. Roy and S. Banerjee, "Behavioral feature engineering for personality prediction using multimodal data," Expert Systems with Applications, vol. 230, pp. 120–135, 2023.
25. R. Mehta and A. Kulkarni, "Explainable multimodal personality prediction using SHAP analysis," IEEE Access, vol. 13, pp. 112233–112250, 2025.

26. S. Verma and R. Patel, “Fairness-aware multimodal personality prediction systems,” IEEE Access, vol. 13, pp. 99876–99895, 2025.
27. A. Singh and M. Tiwari, “Ensemble learning approaches for personality prediction,” Applied Soft Computing, vol. 145, pp. 110–125, 2024.
28. L. Chen et al., “Vision-Mamba: Advanced multimodal personality prediction using transformer architectures,” IEEE Access, vol. 13, pp. 55678–55695, 2025.
29. H. Zhao and Y. Li, “Joint modeling of Big Five and HEXACO personality traits using multimodal learning,” IEEE Access, vol. 13, pp. 77889–77905, 2025.
30. Y. Zhu, X. Shen, and R. Xia, “Personality-aware human-centric multimodal reasoning,” IEEE Trans. Multimedia, vol. 26, pp. 345–359, 2023.
31. S. Gajula and M. Margam, "A Secure and Scalable Cloud-Based Banking Service Model Leveraging AI and Advanced Cyber Security," 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), Houston, TX, USA, 2026, pp. 1-5, doi: 10.1109/ICAIC67076.2026.11395704.
32. N. Verma and P. Shah, “Hybrid feature selection techniques in multimodal systems,” IEEE Access, vol. 12, pp. 22334–22350, 2024.
33. A. Das and R. Roy, “Dimensionality reduction techniques in high-dimensional multimodal data,” IEEE Access, vol. 12, pp. 55667–55685, 2024.
34. M. Patel and S. Mehta, “Cross-modal interaction learning for personality prediction,” IEEE Trans. Artificial Intelligence, vol. 5, no. 2, pp. 145–160, 2025.
35. K. Nair and V. Menon, “Deep feature extraction in multimodal behavioral analysis,” IEEE Access, vol. 13, pp. 66778–66795, 2025.
36. R. Gupta and A. Jain, “Multimodal explainable AI for personality computing,” IEEE Access, vol. 13, pp. 88990–89005, 2025.
37. S. Iyer and P. Kulkarni, “Scalable multimodal AI frameworks for real-world personality prediction,” IEEE Access, vol. 14, pp. 112–130, 2026.