AUTOMATIC PREDICTION OF VIDEO RECORDED JOB INTERVIEW PERFORMANCE BY EXTRACTING FACIAL, AUDIO AND LEXICAL FEATURES

A DISSERTATION SUBMITTED BY

AVINASH KUMAR MISHRA

(18205016)

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF
MASTERS OF TECHNOLOGY IN **COMPUTER SCIENCE AND ENGINEERING**

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING

PUNJAB ENGINEERING COLLEGE, CHANDIGARH

(DEEMED TO BE UNIVERSITY)

CHANDIGARH

2020

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the dissertation entitled **"AUTOMATIC PREDICTION OF VIDEO RECORDED JOB INTERVIEW PERFORMANCE BY EXTRACTING FACIAL, AUDIO AND LEXICAL FEATURES"** in the partial fulfillment of the requirements for the award of the **Master of Technology** in **Computer Science and Engineering** and submitted in the **Department of Computer Science and Engineering** of the Punjab Engineering College, Chandigarh (Deemed to be University), Chandigarh, is an authentic record of my own work carried out during the period from **January'2020** to **June'2020** under the Supervision of **Rupak Agarwal**. The matter presented in this dissertation has not been submitted by me for the award of any other degree of this or any other University/Institute.

Date: 3rd July, 2020 **Avinash Kumar Mishra**

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Rupak Agarwal
Head Talent Bridge Practice
CL Educate Limited

ACKNOWLEDGEMENT

I take this opportunity to express a deep sense of gratitude towards my mentor **Rupak Agarwal**, for providing excellent guidance, encouragement and inspiration throughout this session. Without his valuable guidance, this would never have been a successful one.

I also want to give thanks to all the faculty members of the Department of Computer Science & Engineering, my family and my classmate Sachin Vashistha for their support and encouragement during this time. I want to thank my friends, for making this period such an extraordinary experience in my life.

Last but not least, my heartiest thanks to all, who have been a pillar of support during the time of research and training.

Avinash Kumar Mishra

M.Tech (Computer Science & Engineering)

SID-18205016

ABSTRACT

The landscape of online interviews has changed the scope of hiring best available candidates in less time and with less effort. The aim of this project is **to build a computational framework that can automatically predict the overall rating of a job interview given the audio-visual recordings and score them on various behavioral competencies (both verbal & nonverbal) and then classify them on hireability**. We developed a standardized protocol for collecting structured interview (SI) videos and a set of rubrics for the manual annotation of the recorded videos depending upon the performance against each behavioral attribute on **EvueMe Selection Robot Platform**. Then we extracted **714 Facial Features** tapping micro expressions, head postures and eye-ball movements using an open source library called **OpenFace**. For Audio feature extraction, we used another open source library called **PyAudioAnalysis** and we got **68 prosadic features** related to speaking energy and MFCCs (Mel-Frequency Cepstral Coefficients). We used **Google Cloud Speech API** to generate text from audio and then we used **LIWC 2015** (Linguistic Inquiry & Word Count) to extract **98 features** related to various **psycholinguistic categories** from the transcribed speech of the interview. We also used cutting edge and state of the art **BERT** (Bidirectional Encoder Representation Transformer) **Embeddings** of the transcribed text as an add on to Lexical features apart from LIWC Features. After doing the Exploratory Data Analysis, we then built **separate pipelines for each modality** (Facial, Audio & Lexical) to later combine them for a **Multimodal Analysis.** By analyzing the relative feature weights learned by the **regression models**, our framework will score the pool of candidates and then **classify** them on **hirebility**.

Keywords - Online Interviews, Verbal & Non-Verbal Behaviour Prediction, Facial Features, Prosodic features, Transcribed Speech, BERT, Lexical Features, Regression, Multimodal Analysis

Content

List of Figures

List of Tables

1) Problem Statement

To build a computational framework that can **automatically predict the overall rating of a job interview** given the audio visual recordings and score them on various behavioural competencies like **Communication** & **Confidence** and also classify them on **Hireability** (**Yes/No**).

The steps (objectives) include:

1. Collecting Rated Interview Videos from **EvueMe Selection Robot Platform**
2. Development of a model to automatically extract features from the interview videos. These extracted features will come in 3 buckets.
   - ➔ **Video Features** - Facial expressions mapping, micro expression analysis and eyeball movements
   - ➔ **Prosadic Features** – Extract Voice prosadic features from speec patterns like intonation, pace, emphasis, vocal pitch, voice intensity, pauses, etc
   - ➔ Lexical Features - Analytics of word choice, intent of words, etc
3. With improved human scores and automated multimodal features, build automated scoring systems to predict several behavioural traits like Communication and Confidence and the later classify the pool of candidates into Hireable or Not Hireable.

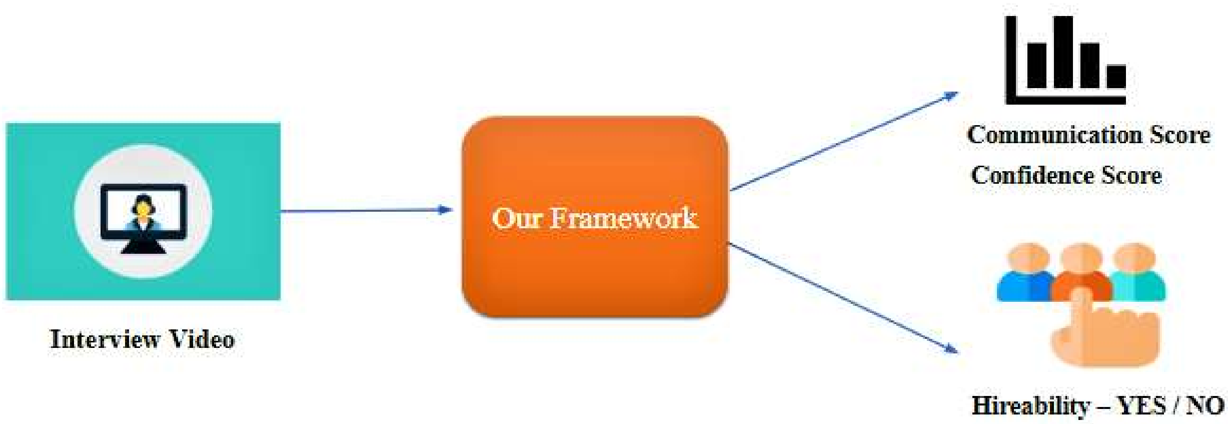

Fig 1. Overall Research Objective

2) Literature Survey

S No.	Title	Algorithm	Year	Research Gap(s)
1	Automated Video Interview Judgment on a Large-Sized Corpus Collected Online	ASR (IBM Bluemax Platform), pyAudioAnalysis, OpenFace, Bag of Words, SVM, RF	2018	No Inter-Rater Agreement, Early Fusion, No Deep Learning Algorithms used
2	Automated Analysis and Prediction of Job Interview Performance	PRAAT toolbox, Shore Framework, OpenFace, LIWC, SVM & Lasso Regression	2016	Unstructured Interviews, Early Fusion, No Deep Learning Algorithms used, Commercial Software used
3	Automated Scoring of Interview Videos using Doc2Vec Multimodal Feature Extraction Paradigm	LIWC, Emotient's FACET SDK, OpenEar, K-Mean Clustering, gensim python toolkit, SVM	2019	No Inter-Rater Agreement, Early Fusion, Commercial Software used
4	Examining Linguistic Content and Skill Impression Structure for Job Interview Analytics in Hospitality	LIWC, OpenFace,OpenEar, Random Forest Regression	2017	Unstructured Interviews, No Inter-Rater Agreement, Early Fusion, No Deep Learning Algorithm(s) used
5	Asynchronous Video Interviews vs. Face-to-Face Interviews For Communication Skill Measurement: A Systematic Study	OPENSmile, PRAAT toolbox, iMotions, VoiceBase online tool, SVM, RF	2018	Early Fusion, Commercial Software used, No Deep Learning Algorithm(s) used
6	Spatiotemporal and Multimodal Analysis of Personality Traits	OpenFace, PyAudioAnalysis, Google Speech API, CNN-GRU, 3D ResNext, LSTNet, RCNN, LSVM	2020	Unstructured Interviews, No Inter-Rater Agreement,

3) Our Proposed Solution

3.1) Data Collection:

All research has shown that past-focused behavioral interview questions, in which the applicant is asked about how he or she has handled work-related situations in the past, yield higher validity coefficients than future-oriented (or hypothetical) behavioral interview questions.

The questions asked were presented one at a time on a computer screen. Participants will be given 1 minute to prepare and up to 2 minutes to respond to each question. A countdown timer was shown on the computer screen. The sequence and timing of the questions was standardized and can not be changed by the interviewee.

Question Asked	**Regression Parameter 1**	**Regression Parameter 2**	**Classification**
Why should we hire you? Where do you see yourself 5 years from now?	Communication (0-100)	Confidence (0-100)	Hireability (Yes/No)

The videos are stored on **Amazon Web Server Platform**. We collected **35** such Manual Rated video interviews (one for each candidate) using **SQL Script** from EvueMe Selection Robot Platform.

3.2) Feature Extraction:

A) Video Features

To obtain measures for facial shape, displayed facial action units (AU), head pose, and gaze, we processed the videos using **OpenFace** (**https://github.com/TadasBaltrusaitis/OpenFace**) as visualized in Fig. below.

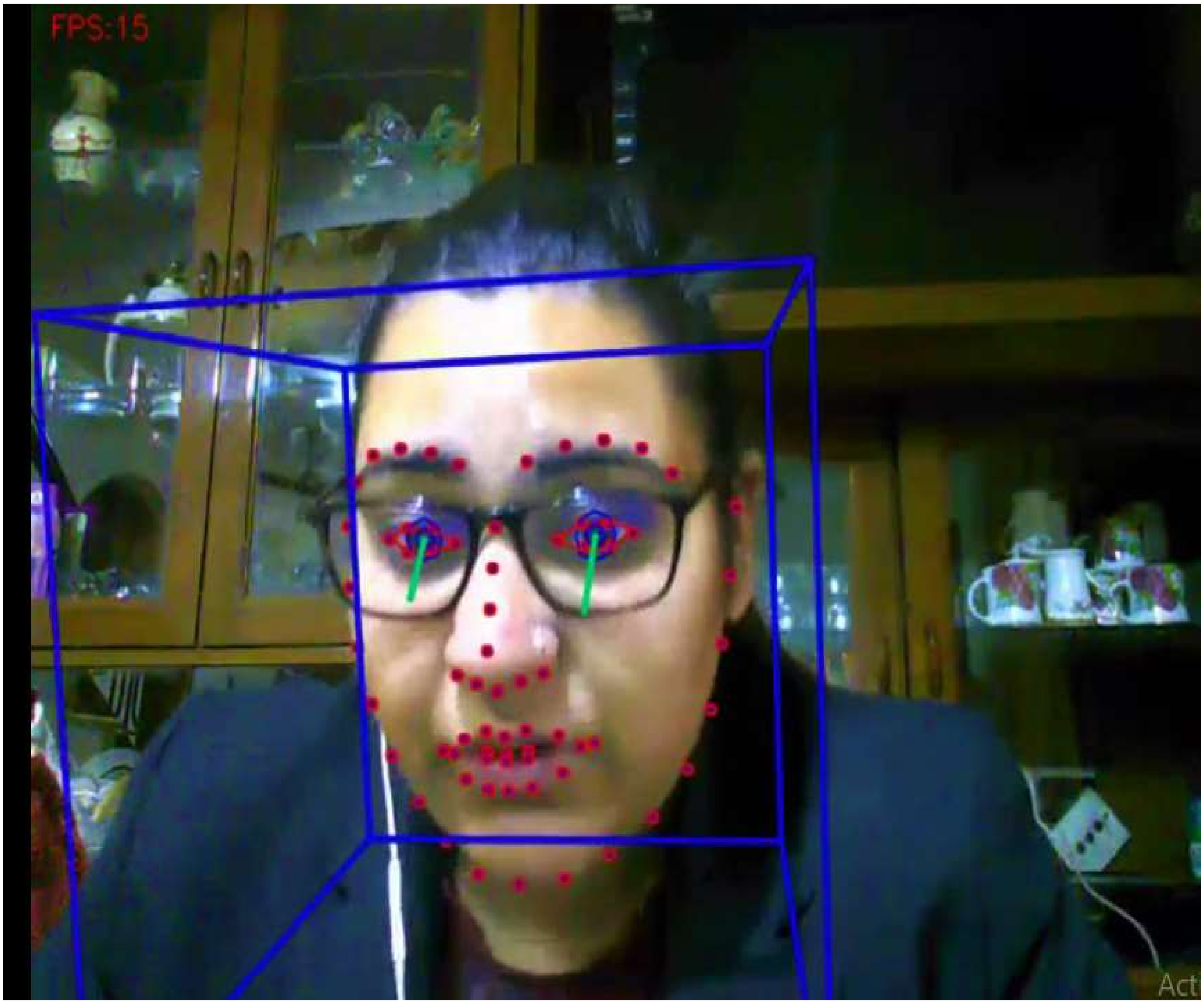

Fig2. Visualization of the facial landmarks, gaze detection and head pose obtained from OpenFace

We get **714 facial features** for each video. However, we get different number of frames for different videos because of their different lengths.

Note: File Size (69585, 714)

<u>List of Facial Features Extracted</u>:

Feature Name	Feature Description	Total Features
gaze_0_x, gaze_0_y, gaze_0_z	Eye gaze direction vector in world coordinates for eye 0 (normalized), eye 0 is the leftmost eye in the image (think of it as a ray going from the left eye in the image in the direction of the eye gaze)	3
gaze_1_x, gaze_1_y, gaze_1_z	Eye gaze direction vector in world coordinates for eye 1 (normalized), eye 1 is the rightmost eye in the image (think of it as a ray going from the right eye in the image in the direction of the eye gaze)	3
gaze_angle_x, gaze_angle_y	Eye gaze direction in radians in world coordinates averaged for both eyes and converted into more easy to use format than gaze vectors. If a person is looking left-right this will results in the change of gaze_angle_x (from positive to negative) and, if a person is looking up-down this will result in change of gaze_angle_y (from negative to positive), if a person is looking straight ahead both of the angles will be close to 0 (within measurement error).	2
eye_lmk_x_0, …....., eye_lmk_y_55	Location of 2D eye region landmarks in pixels. The landmark index can be found below in the same cell.	112
eye_lmk_X_0, ….., eye_lmk_Z_55	Location of 3D eye region landmarks in millimeters. The landmark index can be found below in the same cell.	168

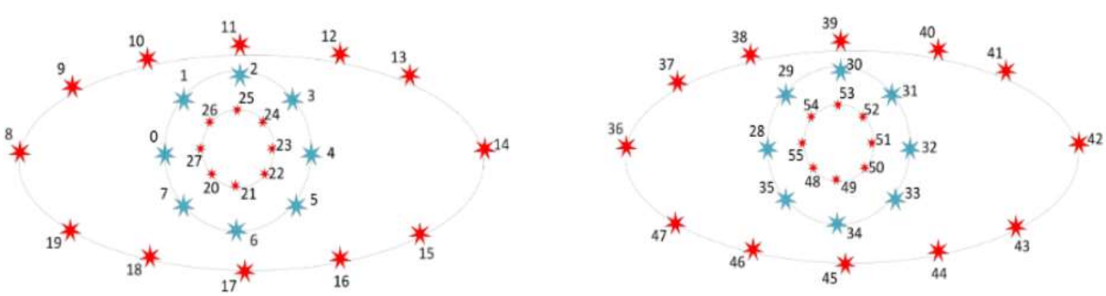

Fig3. Eyeball Movement tracking points

Pose_Tx, Ty, Tz	The location of the head with respect to camera in millimeters (positive Z is away from the camera).	3
Pose_Rx, Ry, Rz	Rotation is in radians around X,Y,Z axes with the convention R = Rx * Ry * Rz, left-handed positive sign. This can be seen as pitch (Rx), yaw (Ry), and roll (Rz).	3
x_0,…...., y_67	Location of 2D landmarks in pixels, the landmark index can be seen below in the same cell.	136
X_0,…....., Z_67	Location of 3D landmarks in millimetres, the landmark index can be seen below.	204

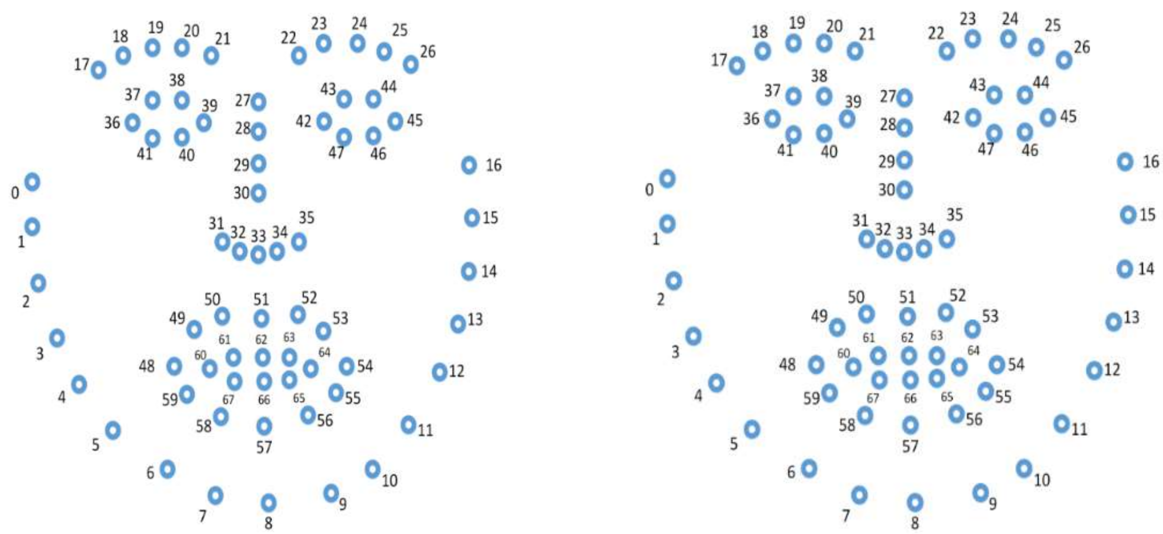

Fig4. Facial Landmarks in 2D (left) & 3D (right)

p_scale, p_rx, ry, rz, tx, ty	Scale, Rotation and Translation terms of the PDM. These are the Parameters of a point distribution model (PDM) that describe the rigid face shape (location, scale and rotation).	6
p_0, ……, p_33	Non-rigid shape parameters. These are the Parameters of a point distribution model (PDM) that describe the non-rigid face shape (deformation due to expression and identity).	34
17 AU's and presence of 18 AU's	Facial AUs are the fundamental actions of individual muscles or groups of muscles to describe human facial expressions over a period of time.	35

Fig5. Some examples of Action Units of FACS (Facial Action Coding System)

In order to describe facial action units, we will use the 18 AU occurrence and 17 AU intensity features provided by OpenFace. While the binary occurrence features indicate the presence of AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU28, and AU45, the intensity features (in the range of [0,5]) represent the intensity of the aforementioned AUs except AU28 (lip sucking).

Table — Facial action unit (AU) features in OpenFace.

AU	Description	AU	Description
AU1	Inner brow raiser	AU14	Dimpler
AU2	Outer brow raiser	AU15	Lip corner depressor
AU4	Brow lowerer	AU17	Chin raiser
AU5	Upper lid raiser	AU20	Lip stretcher
AU6	Cheek raiser	AU23	Lip stretcher
AU7	Lid tightener	AU26	Jaw drop
AU9	Nose wrinkler	AU28	Lip suck
AU10	Upper lip raiser	AU45	Blink
AU12	Lip corner puller		

B) Audio Features

To represent the characteristics of voice, we compute a 34-dimensional feature vector from audio data of videos, including Mel Frequency Cepstral Coefficients (MFCCs), PLP features, Chroma vector (pitch), tonality, energy and entropy related features using **pyAudioAnalysis** framework (an open source Python library). Voice features are extracted for each 50 milliseconds of videos. Consequently, these features form the multivariate time series for describing the voice. Details of the feature extraction process can be found at **https://github.com/tyiannak/pyAudioAnalysis**.

Note- We extracted **68 Audio Features** using it. **File Size (55646,68).**

List of Audio Features Extracted:

Feature Name	Feature Description	Total Features
Zero Crossing Rate	The rate of sign-changes of the signal during the duration of a particular frame.	1
Energy	The sum of squares of the signal values, normalized by the respective frame length.	1
Entropy of Energy	The entropy of sub-frames' normalized energies. It can be interpreted as a measure of abrupt changes.	1
Spectral Centroid	The center of gravity of the spectrum.	1
Spectral Spread	The second central moment of the spectrum.	1
Spectral Entropy	Entropy of the normalized spectral energies for a set of sub-frames.	1
Spectral Flux	The squared difference between the normalized magnitudes of the spectra of the two successive frames.	1
Spectral Rolloff	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.	1

MFCCs	Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.	13
Chroma Vector	A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing).	12
Chroma_Std	Standard Deviation of all 12 Chroma Vectors	1
Delta	local estimate of the derivative of all 34 Short-Term Features like Zero Crossing Rate, Energy etc.	34

C) Lexical Features

Recent studies show that the use of language as an additional modality, improves the performance of estimating personality traits and behavioural attributes. We used state of the art **Google Cloud Speech to Text API** (**https://cloud.google.com/speech-to-text/docs**) which is free to transcribe the audio file of each interview video.

i) LIWC

We used **LIWC 2015** (Linguistic Inquiry & Word Count) software to extract **98** Feature points based upon various psycholinguistic Categories.

Inspired by the efficiency of the conventional method we will use the tool Linguistic Inquiry Word Count (LIWC - **http://liwc.wpengine.com/**) also to extract lexical features related to interview content to analyze the choice and the intent of the words used. The LIWC provides an easy-to-use API that delivers deep insights into people's psychologies and cognitive states by analyzing natural language and text by reporting the counts of various psycholinguistic word categories that include words describing negative emotions (sad, angry, etc.), positive emotions (happy, kind, etc.), different function words (articles, pronouns, etc.), and various content categories (e.g., anxiety, insight) related to psychology, personality, emotion, tone, and more.

<u>List of LIWC Features extracted</u>:

Category	**Feature Name**	**Examples**
Summary Dimensions		
Total Words Count	WC	
Analytical thinking	Analytic	
Clout	Clout	
Authentic	Authentic	
Emotional tone	Tone	
Words Per Sentence	WPS	
Words > 6 letters	Sixltr	
Dictionary words	Dic	
Linguistic Dimensions		
Total function words (row no =13 to 25)	funct	it, to, no, very
Total pronouns (row = 14 to 19)	pronoun	I, them, itself
Personal pronouns (row = 15 to 19)	ppron	I, them, her
1st person singular	i	I, me, mine
1st person plural	we	we, us, our
3rd person singular	shehe	she, her, him
3rd person plural	they	they, their, they'd
Impersonal pronouns	ipron	it, it's, those
Articles	article	a, an, the
Prepositions	prep	to, with, above
Auxiliary Verbs	auxverb	am, will, have
Common Adverbs	adverb	very, really
Conjunctions	conj	and, but, whereas

Negations	negate	no, not, never
Other Grammar		
Common Verbs	verb	eat, come, carry
Common Adjectives	adj	free, happy, long
Comparisons	compare	greater, best, after
Interrogatives	interrog	how, when, what
Numbers	number	second, thousand
Quantifiers	quant	few, many, much
Psychological Processes		
Affect (row no = 35 to 39)	affect	happy, cried
Positive Emotion	posemo	love, nice, sweet
Negative Emotion (row no = 37 to 39)	negemo	hurt, ugly, nasty
Anxiety	anx	worried, fearful
Anger	anger	hate, kill, annoyed
Sadness	sad	crying, grief, sad
Social		
Social (row no = 42 to 45)	social	mate, talk, they
Family	family	daughter, dad, aunt
Friends	friend	buddy, neighbor
Female references	female	girl, her, mom
Male references	male	boy, his, dad
Cognitive Process		
Cognitive processes (row no = 48 to 53)	cogproc	cause, know, ought
Insight	insight	think, know
Causation	cause	because, effect
Discrepancy	discrep	should, would

Tentative	tentat	maybe, perhaps
Certainty	certain	always, never
Differentiation	differ	hasn’t, but, else
Perceptual Processes		
Perceptual Processes (row no = 56 to 58)	percept	look, heard, feeling
See	see	view, saw, seen
Hear	hear	listen, hearing
Feel	feel	feels, touch
Biological Processes		
Biological processes (row no = 61 to 64)	bio	eat, blood, pain
Body	body	cheek, hands, spit
Health	health	clinic, flu, pill
Sexual	sexual	horny, love, incest
Ingestion	ingest	dish, eat, pizza
Drives		
Drives (row no = 67 to 71)	drives	friend, danger, bully, win
Affiliation	affiliation	ally, friend, social
Achievement	achieve	win, success, better
Power	power	superior, bully
Reward	reward	take, prize, benefit
Risk	risk	danger, doubt
Time orientations (TimeOrient)		
Past Focus	focuspast	ago, did, talked
Present Focus	focuspresent	today, is, now
Future Focus	focusfuture	may, will, soon
Relativity		

Relativity (row no = 78 to 80)	relativ	area, bend, exit
Motion	motion	arrive, car, go
Space	space	down, in, thin
Time	time	end, until, season
Personal Concerns		
Work	work	job, majors, xerox
Leisure	leisure	cook, chat, movie
Home	home	kitchen, landlord
Money	money	audit, cash, owe
Religion	relig	altar, church
Death	death	bury, coffin, kill
Informal Language		
Informal Language (row no = 90 to 94)	informal	btw, damn, Imean, hm
Swear words	swear	fuck, damn, shit
Netspeak	netspeak	btw, lol, thx
Assent	assent	agree, OK, yes
Nonfluencies	nonflu	er, hm, umm, ahh
Fillers	filler	Imean, youknow, like
Punctuation Marks		
Period	Period	
Comma	Comma	
Colon	Colon	
Semi Colon	SemiC	
Question Mark	QMark	
Excalmation Mark	Exclam	
Dash	Dash	

Quotation	Quote	
Apostrophe	Apostro	
Parenthesis	Parenth	
Other Punctuation Marks	OtherP	

ii) BERT Embeddings

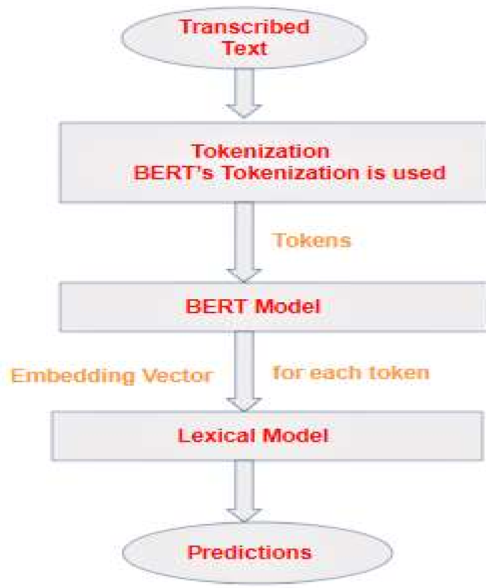

Fig6. BERT Embeddings from Transcribed Speech

To make our model generalizable, we employed pytorch-transformers' implementation of pretrained BERT model (**https://github.com/huggingface/transformers**) to generate embeddings for each token in the transcripts. The BERT model was applied to the whole transcript of each video, separately. BERT model infers the embeddings of each word in the transcript considering its **context**. Flow of our transcribed speech model can be seen in Fig.below. BERT embeddings were used as input features to our model along with LIWC features.

3.3) Exploratory Data Analysis

i) Facial Correlation Analysis & Feature Selection

After finding the correlation matrix of all 714 features, we selected **58** features whose **correlation coefficient** were greater than **0.6** out of which

- ➔ 56 are our Audio features and 2 are for judging criteria in case of **Regression**
- ➔ 57 are our Audio features and 1 is for judging criteria in case of **Classification**

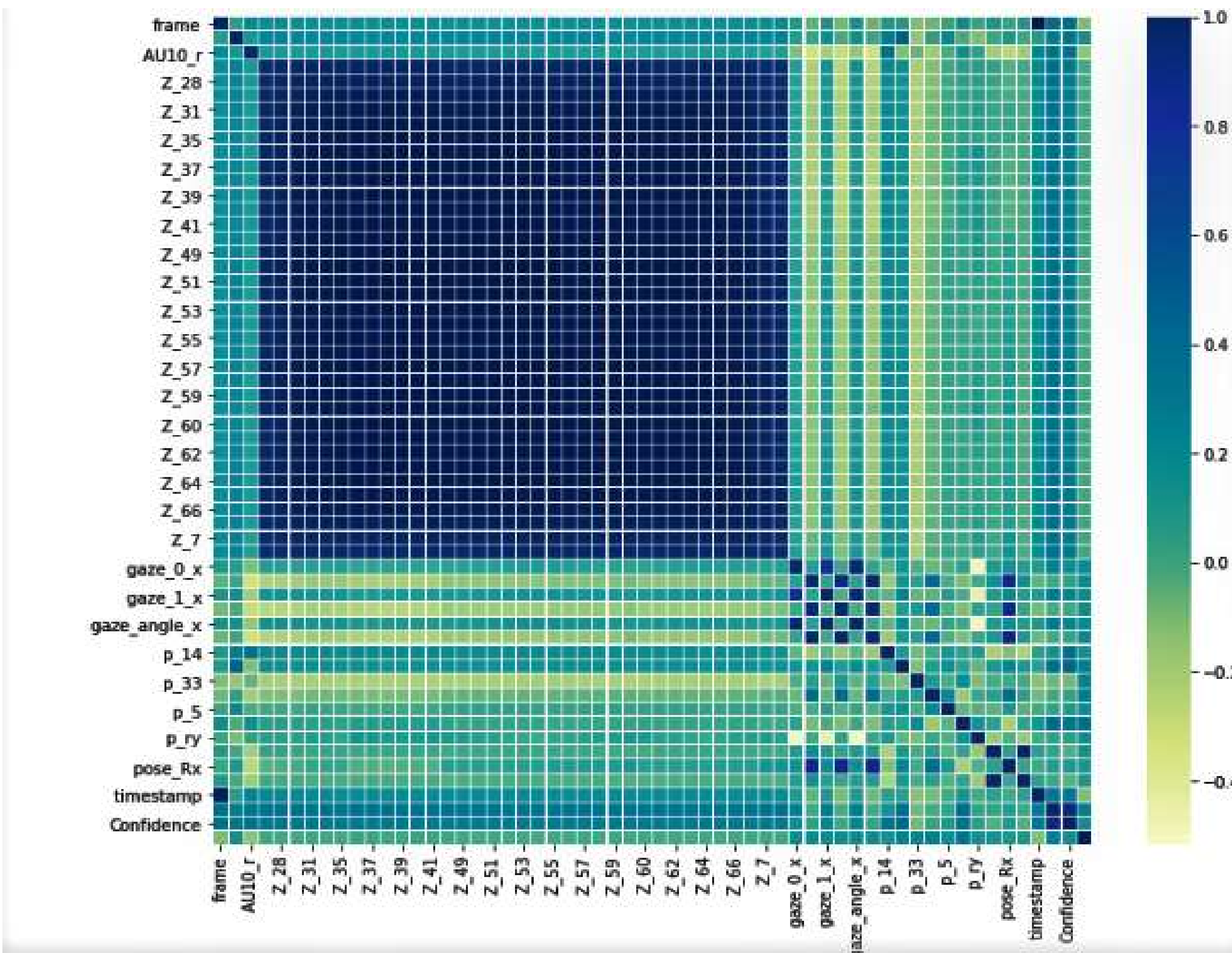

Fig7. Correlation Matrix of selected Facial Feature

58 such features are -

Z_6 Z_27 Z_28 Z_29 Z_31 Z_32 Z_35 Z_36 Z_37 Z_38 Z_39 Z_40 Z_41 Z_48 Z_49 Z_50 Z_51 Z_52 Z_53 Z_54 Z_55 Z_56 Z_57 Z_58 Z_59 Z_60 Z_61 Z_62 Z_63 Z_64 Z_65 Z_66 Z_67 Z_7 Z_8

gaze_0_x gaze_0_y gaze_1_x gaze_1_y gaze_angle_x gaze_angle_y

p_14 p_3 p_33 p_4 p_5 p_7 p_ry p_rz pose_Rx pose_Rz

timestamp Communication Confidence Hireability frame AU07_r AU10_r

ii) Audio Correlation Analysis & Feature Selection

After finding the correlation matrix of all 68 features, we selected **15** features whose **correlation coefficient** were greater than **0.6** out of which

- ➔ 13 are our Audio features and 2 are for judging criteria in case of Regression
- ➔ 14 are our Audio features and 1 is for judging criteria in case of Classification

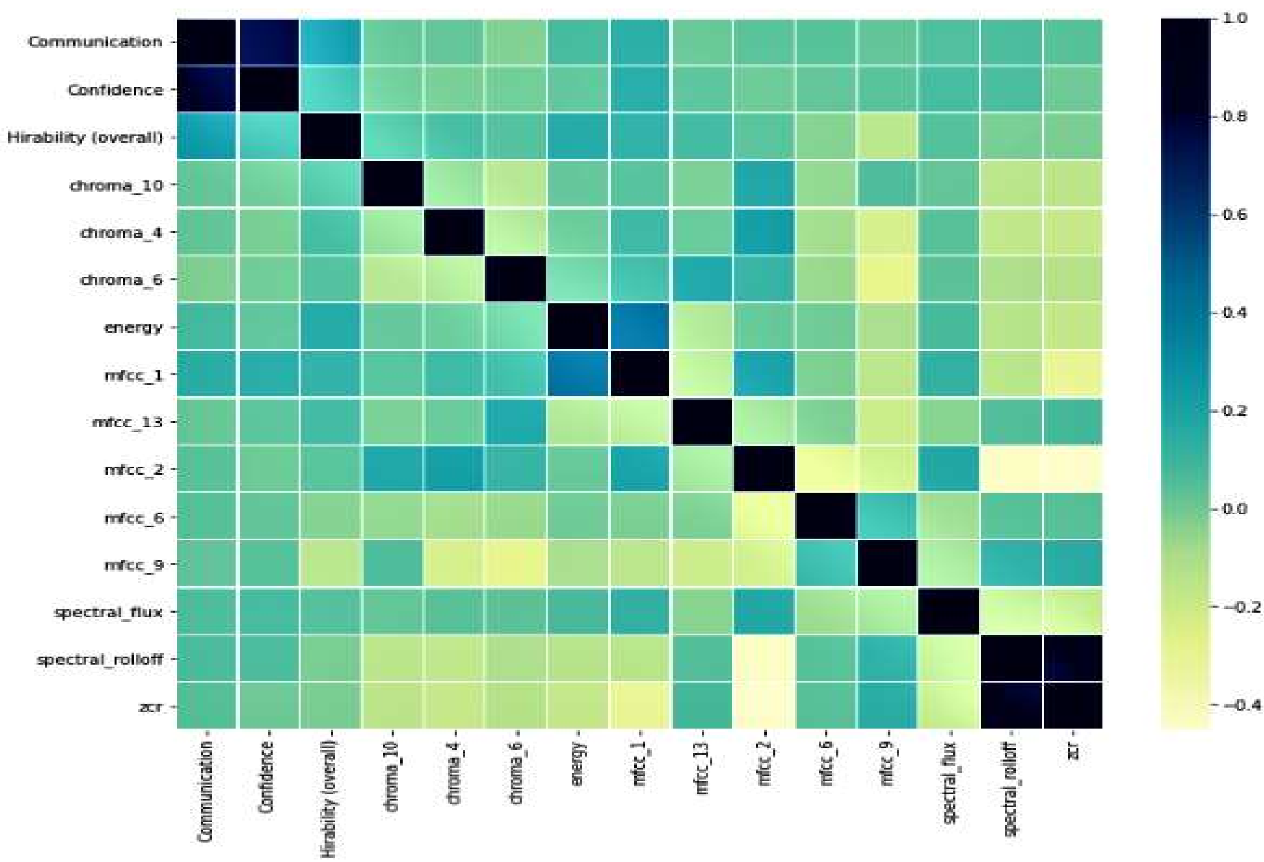

Fig8. Correlation Matrix of selected Audio Features

15 such features are:

chroma_10	mfcc_1	mfcc13	Communication
chroma_4	mfcc_2	spectra_flux	Confidence
chroma_6	mfcc_6	spectral_rolloff	Hireability
Energy	mfcc_9	zcr	

iii) Lexical (LIWC) Correlation Analysis & Feature Selection

After finding the correlation matrix of all 98 LIWC features, we selected **24** features whose **correlation coefficient** were greater than **0.6** out of which

- ➔ 22 are our Audio features and 2 are for judging criteria in case of Regression
- ➔ 23 are our Audio features and 1 is for judging criteria in case of Classification

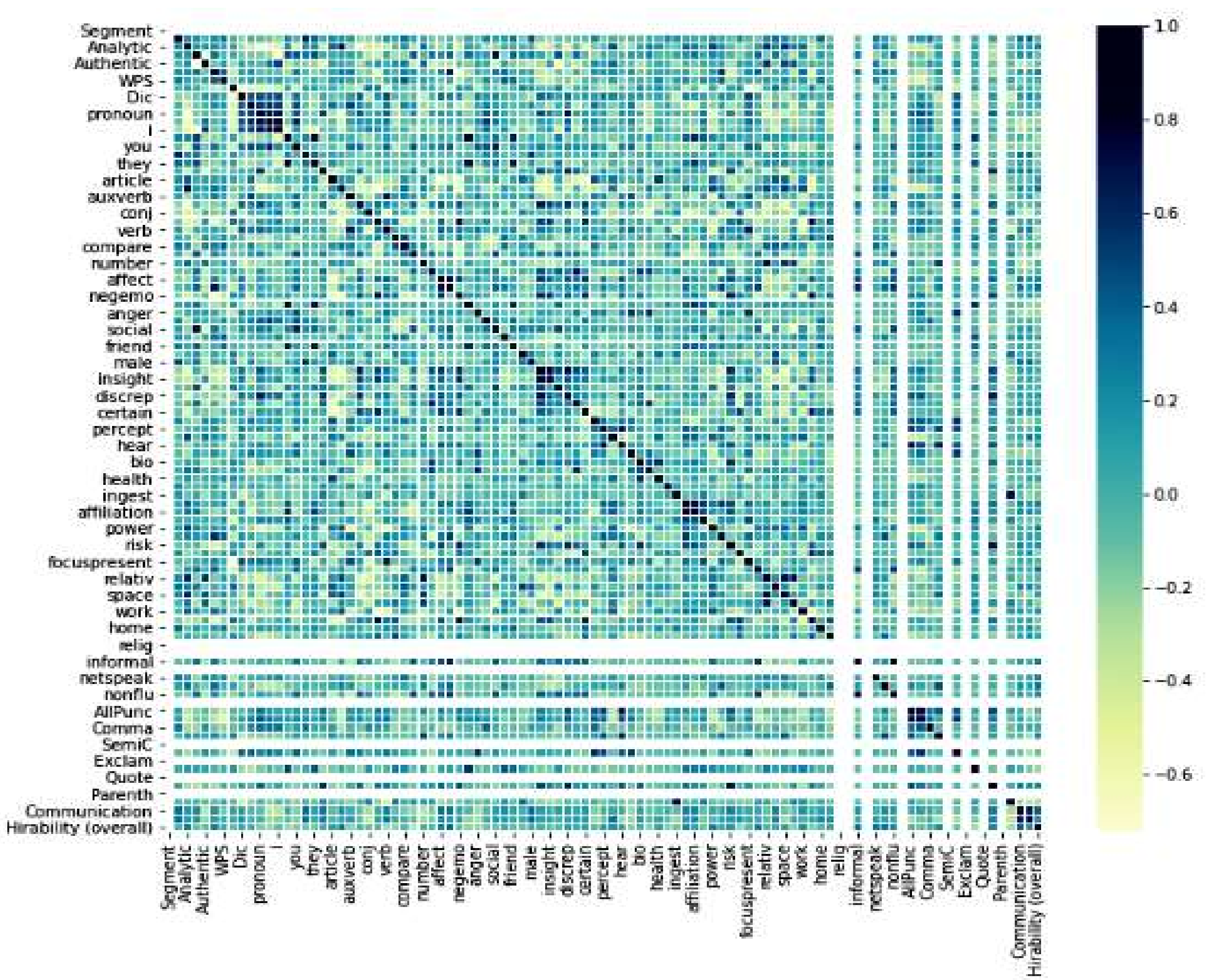

Fig9. Correlation Matrix of LIWC Features

3.4) Model building

Train Test Split - 21 Videos in Train and **10** Videos for Test were taken by making sure that if a candidate is in Train Data, all frames of his/her video are in Test Data. We also applied **Standardization** (Mean = 0 & Standard Deviation =1) to our dataset.

i) Facial Model

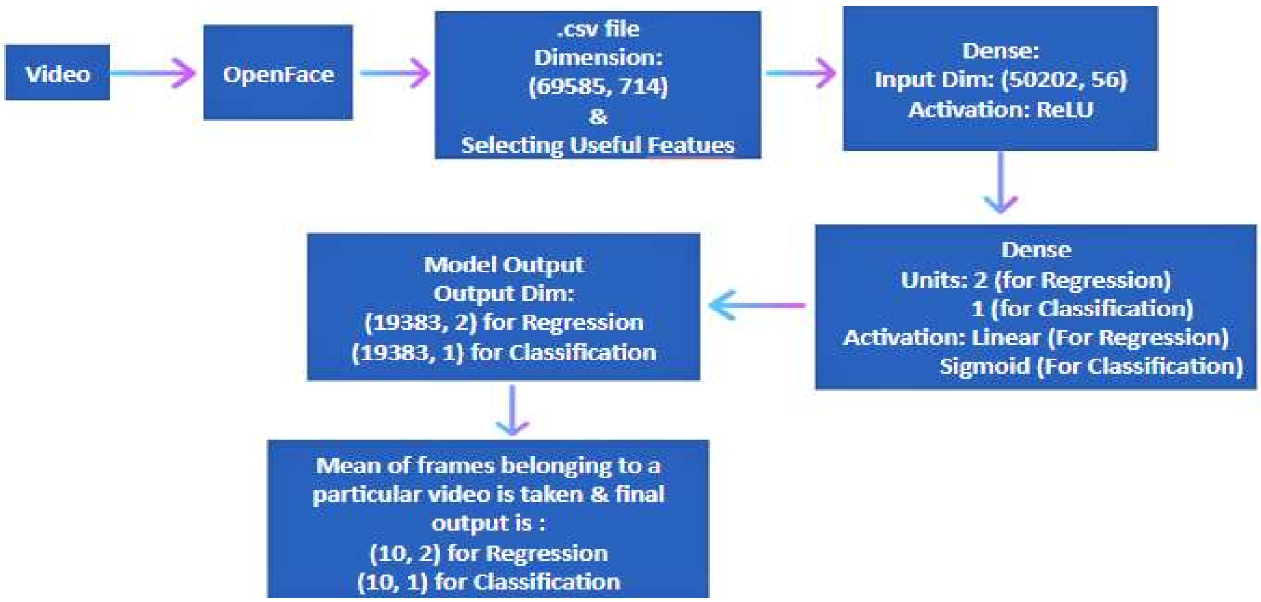

Fig10. Facial Pipeline

ii) Audio Model

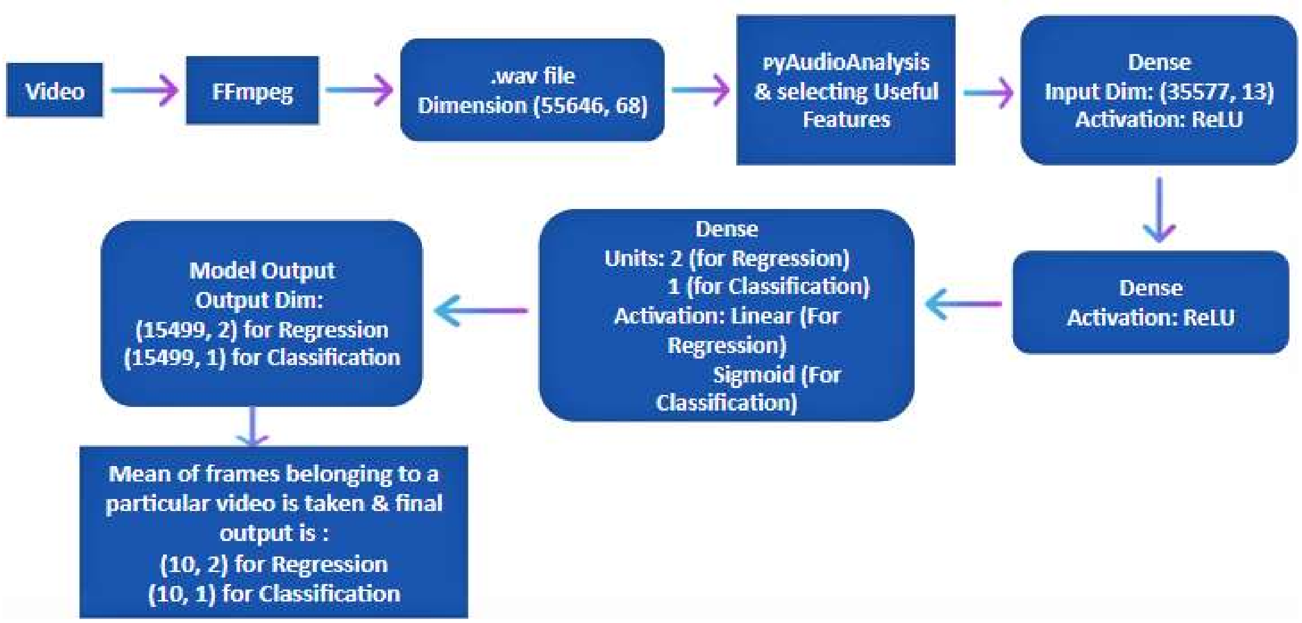

Fig11. Audio Pipeline

iii) Lexical Model

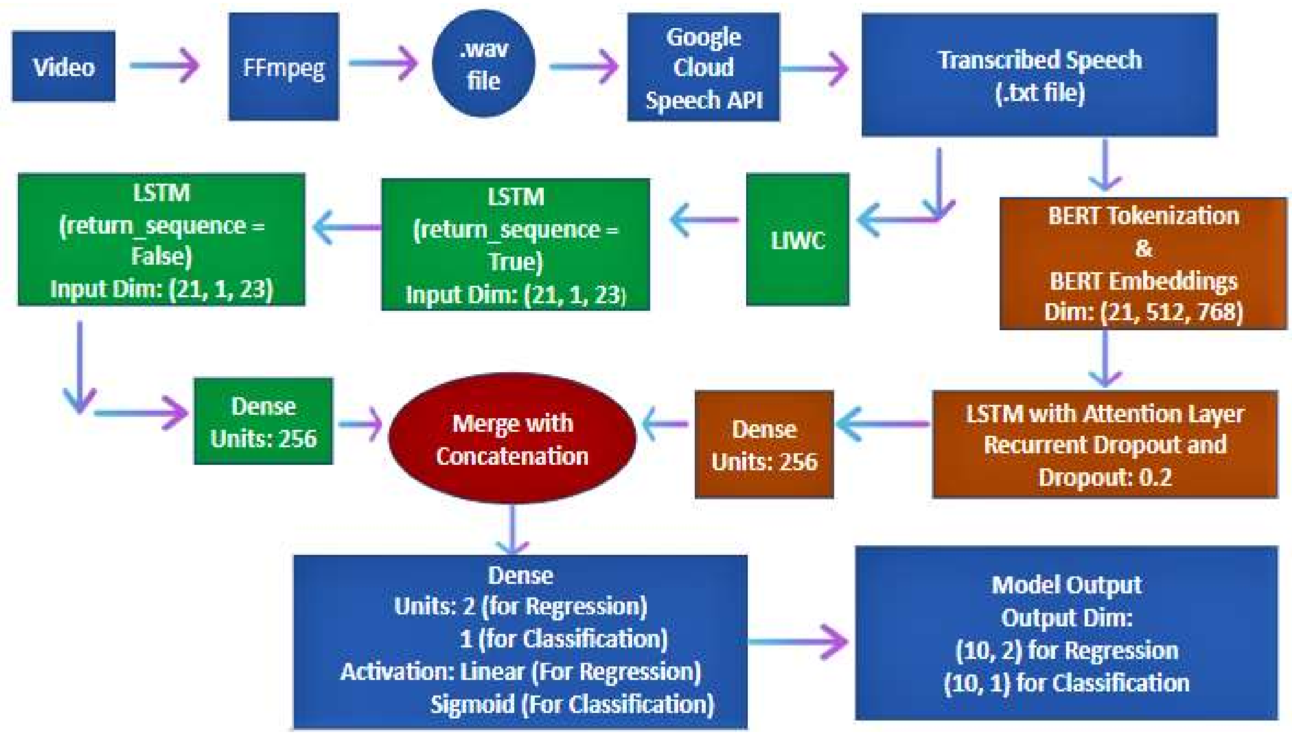

Fig12. Lexical Pipeline

iv) Final Combined Model

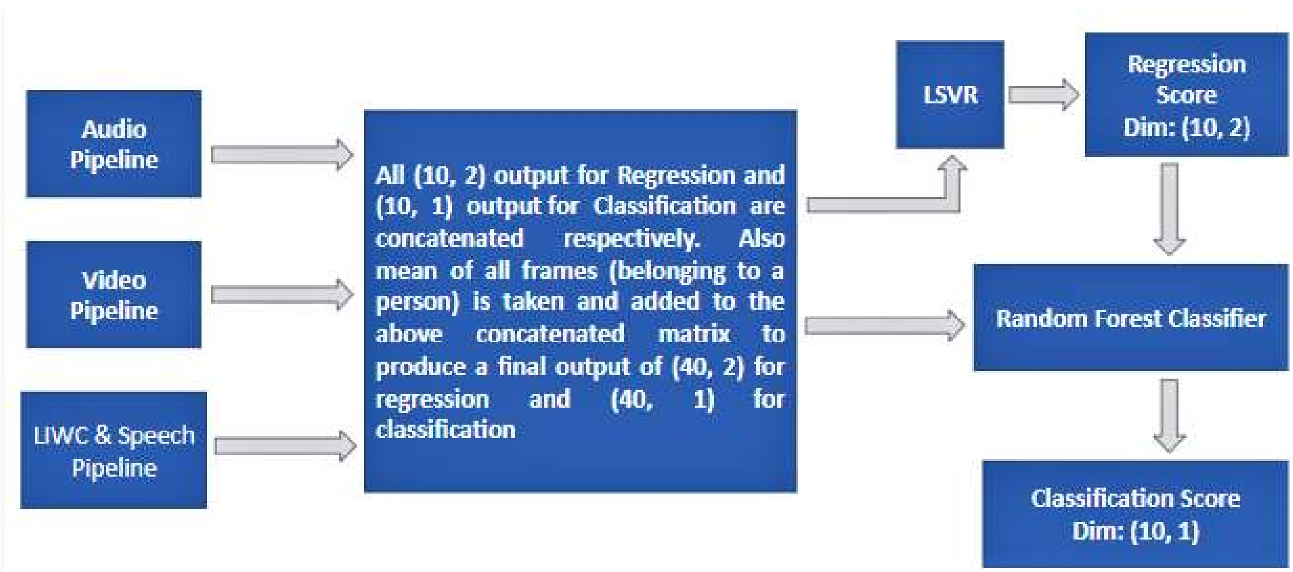

Fig13. Multimodal Analysis

v) Results

➔ Regression Score on Behavioural Attributes

	Name	Communication Actual	Communication Predcited	Confidence Actual	Confidence Predicted
0	Anshuman	60.0	64.214368	65.0	68.179111
1	Gulshan	70.0	67.880970	75.0	70.801656
2	Himanshu	70.0	67.071783	75.0	71.307684
3	Kushal	60.0	67.909946	70.0	70.662648
4	Mohit	85.0	66.384471	85.0	71.162922
5	Rajat	65.0	66.749857	70.0	71.233612
6	Sasikumar	70.0	67.748747	65.0	71.145924
7	Shikhar	75.0	66.813856	80.0	71.263665
8	Shivani	75.0	67.925974	80.0	70.668727
9	Vineesha	75.0	67.059757	75.0	71.315636

➔ Classification on Hireability

	Name	Actual Hirability	Predicted Hirability
0	Anshuman	1	0
1	Gulshan	0	0
2	Himanshu	0	0
3	Kushal	0	0
4	Mohit	1	0
5	Rajat	0	0
6	Sasikumar	0	0
7	Shikhar	0	0
8	Shivani	0	0
9	Vineesha	1	0

4) References

1) https://ieeexplore.ieee.org/document/8925439 (Slices of Attention in Asynchronous Video Job Interviews)

2) https://ieeexplore.ieee.org/document/8660507 (TensorFlow-based Automatic Personality Recognition Used in Asynchronous Video Interviews)

3)https://www.researchgate.net/publication/309493007_Automated_Scoring_of_Interview_Videos_using_Doc2Vec_Multimodal_Feature_Extraction_Paradigm

4) https://www.idiap.ch/~gatica/publications/MuralidharGatica-mum17.pdf (Examining Linguistic Content and Skill Impression Structure for Job Interview Analytics in Hospitality)

5) https://www.researchgate.net/publication/279955258_Computational_Analysis_of_Human-Robot_Interactions_through_First-Person_Vision_Personality_and_Interaction_Experience

6) https://ieeexplore.ieee.org/document/6746207 (Hire me: Computational Inference of Hireability in Employment Interviews Based on Nonverbal Behavior)

7) https://ieeexplore.ieee.org/document/7829933 (Scaling personality traits of interviewees in an online job interview by vocal spectrum and facial cue analysis)

8) https://ieeexplore.ieee.org/document/7163127 (Automated prediction and analysis of job interview performance: The role of what you say and how you say it)

9) https://ieeexplore.ieee.org/document/8273646 (Automated video interview judgment on a large-sized corpus collected online)

10)https://www.researchgate.net/publication/307889146_Automatic_Scoring_of_Monologue_Video_Interviews_Using_Multimodal_Cues

11) https://www.researchgate.net/publication/309613999_Asynchronous_video_interviews_vs_face-to-face_interviews_for_communication_skill_measurement_a_systematic_study

12)https://www.idiap.ch/~gatica/publications/MuralidharNguyenFrauendorferOdobezSchmidGatica-icmi16.pdf (Training on the Job: Behavioral Analysis of Job Interviews in Hospitality)

13) http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.5734&rep=rep1&type=pdf (Towards Automatic Scoring of Non-Native Spontaneous Speech)

14) https://www.idiap.ch/~gatica/publications/NguyenGatica-tmm16.pdf (Hireability in the wild: Analysis of online conversational video resumes)

15) https://arxiv.org/pdf/1609.06773.pdf (Joint CTC-Attention based end-to-end Speech Recognition using Multi-Task Learning)

16) https://www.researchgate.net/publication/336782624_Multimodal_Fusion_of_BERT-CNN_and_Gated_CNN_Representations_for_Depression_Detection

17) https://arxiv.org/abs/1702.02510 (AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups)

18) https://arxiv.org/abs/1908.02265 (ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks)

19)https://arxiv.org/pdf/1904.09078.pdf (EmbraceNet: A robust deep learning architecture for multi-classification)

20)https://arxiv.org/pdf/1908.05787.pdf (M-BERT: Injecting Multimodal Information in the BERT Structure)

21)https://arxiv.org/pdf/1804.08046.pdf (First Impressions: A Survey on Vision-based Apparent Personality Trait Analysis)

22)http://www.ijircce.com/upload/2017/june/231_Behavioral%20Analysis%20of%20Candidates%20u %20Sentiment%20Analysis%20and%20Emotion%20Mining%20for%20Recruitment.pdf(Behaviora Analysis of Candidates using Sentiment Analysis and Emotion Mining for Recruitment)

23) https://www.researchgate.net/publication/257652062_A_Job_Interview_Simulation_Social_Cue-Based_Interaction_with_a_Virtual_Character (A Job Interview Simulation: Social Cue-Based Interaction with a Virtual Character)

24) https://ieeexplore.ieee.org/abstract/document/8728458/ (An Automated System for the Assessment of Interview Performance through Audio & Emotion Cues)

25) https://sigport.org/documents/attention-based-dialog-state-tracking-conversational-interview-coaching (Attention-based Dialog State Tracking For Conversational Interview Coaching)

26) https://arxiv.org/pdf/1804.08348.pdf (Deep Facial Expression Recognition: A Survey)

27) https://arxiv.org/abs/1710.10198.pdf (Detection and Analysis of Human Emotions through Voice and Speech Pattern Processing)

28) https://www.intechopen.com/books/intelligent-video-surveillance/detecting-micro-expressions-in-real-time-using-high-speed-video-sequences (Detecting Micro-Expressions in Real Time Using High Speed Video Sequences)

29) https://ieeexplore.ieee.org/document/8373897 (Facial Micro-expressions Grand Challenge 2018: Evaluating spatio-temporal features for classification of objective classes)

30) https://ieeexplore.ieee.org/document/8663219 (Fraud Detection by Facial Expression Analysis Using Intel Real Sense and Augmented Reality)

31) https://arxiv.org/pdf/1806.05781.pdf (A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges)

32) https://s3-us-west-2.amazonaws.com/downloads.liwc.net/LIWC2015_OperatorManual.pdf (Linguistic Inquiry and Word Count: LIWC2015)

33) https://ieeexplore.ieee.org/document/8883064 (Micro and macro facial expression recognition using advanced Local Motion Patterns)

34) https://ieeexplore.ieee.org/document/8424834 (Integrating Vision and Language for First-Impression Personality Analysis)

35) https://publications.idiap.ch/downloads/papers/2013/Nguyen_ACMICMI_2013.pdf (Multimodal Analysis of Body Communication Cues in Employment Interviews)

36) https://ieeexplore.ieee.org/document/8819708 (Natural Language Processing based Jaro-The Interviewing Chatbot)

37) https://ieeexplore.ieee.org/document/7829933 (Scaling Personality Traits of Interviewees in an Online Job Interview by Vocal Spectrum and Facial Cue Analysis)

38) https://ieeexplore.ieee.org/document/8769559 (Sentiment Classification and Prediction of Job Interview Performance)

39) https://arxiv.org/pdf/1801.07481.pdf (Survey on Emotional Body Gesture Recognition)

40) https://www.researchgate.net/publication/327550280_Video_Analysis_for_Micro-_Expression_Spotting_and_Recognition (Video Analysis for Micro- Expression Spotting and Recognition)

41) https://arxiv.org/pdf/1904.01766.pdf (VideoBERT: A Joint Model for Video and Language Representation Learning)

42)https://www.researchgate.net/publication/318848763_Which_Prosodic_Features_Contribute_to_the_Recognition_of_Dramatic_Attitudes (Which Prosodic Features Contribute to the Recognition of Dramatic Attitudes?)

43) https://arxiv.org/pdf/1703.07015.pdf (Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks)

44) https://arxiv.org/pdf/1703.07015.pdf (Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks)

45) https://arxiv.org/pdf/1908.05787.pdf (M-BERT: Injecting Multimodal Information in the BERT Structure)

46) https://arxiv.org/pdf/1906.00295.pdf {https://github.com/yaohungt/Multimodal-Transformer} (Multimodal Transformer for Unaligned Multimodal Language Sequences)

47)https://www.researchgate.net/publication/340548401_Spatiotemporal_and_Multimodal_Analysis_of_Personality_Traits (Spatiotemporal and Multimodal Analysis of Personality Traits)