Intelligent Systems Reference Library 206

Mayuri Mehta
Philippe Fournier-Viger
Maulika Patel
Jerry Chun-Wei Lin *Editors*

Tracking and Preventing Diseases with Artificial Intelligence

Springer

Intelligent Systems Reference Library

Volume 206

Series Editors

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

Lakhmi C. Jain, KES International, Shoreham-by-Sea, UK

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included. The list of topics spans all the areas of modern intelligent systems such as: Ambient intelligence, Computational intelligence, Social intelligence, Computational neuroscience, Artificial life, Virtual society, Cognitive systems, DNA and immunity-based systems, e-Learning and teaching, Human-centred computing and Machine ethics, Intelligent control, Intelligent data analysis, Knowledge-based paradigms, Knowledge management, Intelligent agents, Intelligent decision making, Intelligent network security, Interactive entertainment, Learning paradigms, Recommender systems, Robotics and Mechatronics including human-machine teaming, Self-organizing and adaptive systems, Soft computing including Neural systems, Fuzzy systems, Evolutionary computing and the Fusion of these paradigms, Perception and Vision, Web intelligence and Multimedia.

Indexed by SCOPUS, DBLP, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/8578

Mayuri Mehta · Philippe Fournier-Viger ·
Maulika Patel · Jerry Chun-Wei Lin
Editors

Tracking and Preventing Diseases with Artificial Intelligence

Springer

Editors
Mayuri Mehta
Department of Computer Engineering
Sarvajanik College of Engineering
and Technology
Surat, Gujarat, India

Maulika Patel
Department of Computer Engineering
G. H. Patel College of Engineering
and Technology
Charutar Vidya Mandal University
Vallabh Vidyanagar, Gujarat, India

Philippe Fournier-Viger
School of Humanities and Social Sciences
Harbin Institute of Technology (Shenzhen)
Shenzhen, Guangdong, China

Jerry Chun-Wei Lin
Department of Computer Science,
Electrical Engineering and
Mathematical Sciences
Western Norway University of
Applied Sciences
Bergen, Norway

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-030-76731-0 ISBN 978-3-030-76732-7 (eBook)
https://doi.org/10.1007/978-3-030-76732-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

All around the world, the spread of infectious diseases is a major concern as it directly impacts the health of people. While some infectious diseases may have a minor impact on society, some can have major impacts such as the recent SARS-CoV-2 coronavirus pandemic, also known as COVID-19.

To cope with the spread of infectious diseases, some traditional approaches are used such as to study the effect of medicines and develop new ones, design appropriate vaccines, and enforce various measures such as washing hands, wearing face masks, and doing temperature checks. However, despite the usage of such measures and medical advancements, there remain several incurable diseases for which prevention is the only cure. Besides, time is often critical when coping with new diseases that are highly contagious such as COVID-19 as no vaccine or very effective medicine is initially available.

To cope with these challenges, artificial intelligence (AI) has been rapidly adopted to assist physicians in diagnosis, disease tracking, prevention, and control. Due to increasing the availability of electronic healthcare data and rapid progress of analytics techniques, a lot of research is being carried out in this area by applying machine learning and data mining techniques to assist the medical professionals for making a preliminary evaluation.

This book is a collection of 11 chapters that provides a timely overview of recent advances in this area, that is, to use artificial intelligence techniques for tracking and preventing diseases. The target audience of this book is researchers, practitioners, and students. A brief description of each chapter is given below.

In Chap. 1, four approaches to identify stress by recognizing the emotional state of a person have been proposed. Pradeep et al. have analyzed the performance of the proposed approaches using Surrey Audio-Visual Expressed Emotion (SAVEE) and ENTERFACE databases. The results illustrate the considerable reduction in computational time and show that vector quantization-based features perform better than mel-frequency cepstral coefficients feature.

In Chap. 2, Fayemiwo et al. compared various approaches for the detection of COVID-19 from X-ray images. The problem is viewed as a classification problem with two classes (normal vs COVID-19) or three classes (normal, pneumonia, and COVID-19). A fine-tuned VGG-19 convolutional neural network with deep transfer learning shows that high accuracy can be obtained (from 89% to 99% depending on the scenario.

In Chap. 3, Falguni et al. aim to develop an intelligent diagnostic system for glaucoma—an eye-related disease, from the data obtained through clinicians by various examination devices or equipment used in ophthalmology. The classification is done by using a hybrid approach using artificial neural network, Naïve Bayes algorithms, decision tree algorithms, and 18 medical examination parameters for a patient. FGLAUC-99 is developed with J48, Naïve Bayes, and MLP classifiers with accuracy of 99.18%. The accuracy is not compared with other classifiers as the dataset is exclusively developed.

In Chap. 4, Pathak et al. have introduced two approaches, one based on a simple neural network and another based on a deep convolutional neural network, for diagnosis of tuberculosis disease. To evaluate the performance of the proposed approaches, they conducted experiments using tuberculosis chest X-ray dataset available on Kaggle and received classification accuracy of 99.24%.

In Chap. 5, Sarumi and Leung proposed an adaptive Naive Bayes-based machine learning algorithm for efficient prediction of genes in the genome of eukaryotic organisms. The adaptive Naive Bayes algorithm provided a sensitivity, specificity, and accuracy of 81.52%, 94.01%, and 96.02%, respectively, on discovering the protein-coding genes from the human genome chromosome GRCh37.

In Chap. 6, Deshpande et al. presented a survey work on different areas where microscopic imaging of blood cells is used for disease detection. A small note on blood composition is first discussed, which is followed by a generalized methodology for microscopic blood image analysis for certain application of medical imaging. Several models using microscopic blood cell image analysis are also summarized for disease detection.

In Chap. 7, Mahajan and Rana presented a comprehensive review of the recent clinical named entity classification using rule-based, deep learning-based, and hybrid approaches. The efficacy of clinical named entity recognition (NER) techniques for information extraction is also discussed and several experiments are then evaluated to show the state-of-the-art models with high accuracy by combining deep learning (DL) models with a sequential model.

In Chap. 8, the topic of disease diagnosis from CT scan images is discussed. Sajja et al. present a generic and hybrid intelligent architecture for disease diagnosis. The architecture can classify images into various disease categories using a convolutional neural network and is applied for detecting the COVID-19 disease. The design of the model is presented in detail with an experimental evaluation and a discussion of applications for other disease diagnoses using radiology images, as well as possibilities for future work.

In Chap. 9, skin lesion classification problem is addressed. Rock et al. developed an online system to assist doctors to quickly diagnose skin disease through skin lesion observation. Results demonstrate 78% testing accuracy and 84% training and validation accuracy.

In Chap. 10, Oza et al. have discussed various mammogram classification techniques that are categorized based on function, probability, rule, and similarity. They have presented comparative analysis of these techniques including strengths, drawbacks, and challenges. A few mechanisms to deal with these challenges have been described. In addition, some publicly available mammogram datasets are discussed in this chapter.

In Chap. 11, Sachdev et al. have presented the state-of-the-art similarity-based and feature-based chemogenomic techniques for the prediction of interaction between drug compounds and proteins. They have illustrated comparison of these techniques including their merits and demerits.

Surat, India — Mayuri Mehta
Shenzhen, China — Philippe Fournier-Viger
Vallabh Vidyanagar, India — Maulika Patel
Bergen, Norway — Jerry Chun-Wei Lin
March 2021

Contents

Contributors

Rajanikanth Aluvalu Department of CSE, Vardhaman College of Engineering, Hyderabad, India

C. Alvino Rock Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

Samson A. Arekete Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

C. Arvinthan Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

E. Bijolin Edwin Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

A. D. Darji Department of Electronics Engineering, Sardar Vallabhbhai Patel National Institute of Technology, Surat, Gujrat, India;
S.V.N.I.T., Ichhanath, Surat, Gujarat, India

Nilkanth Mukund Deshpande Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Lavale, Pune, India;
Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar, India;
Symbiosis International (Deemed University), Pune, India

Michael A. Fayemiwo Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Shilpa Shailesh Gite Department of Computer Science, Symbiosis Institute of Technology, Symbiosis Centre for Applied AI (SCAAI), Lavale, Pune, India;
Symbiosis International (Deemed University), Pune, India

Manoj K. Gupta Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Katra, India

Richard Jayaraj Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

R. J. S. Jeba Kumar Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

B. Kevin Joseph Paul Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

Swathi S. Kundaram S.C.E.T., Athwalines, Surat, Gujarat, India

Carson K. Leung University of Manitoba, Winnipeg, MB, Canada

Pranita Mahajan SIESGST, Navi Mumbai, India;
SVNIT, Surat, India

Mba O. Odim Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Adewale O. Ogunde Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Bosede O. Oguntunde Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Theresa O. Ojewumi Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Oluwabunmi O. Olaniyan Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Toluwase A. Olowookere Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Idowu S. Oyetade Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

Parita Oza Nirma University, Ahmedabad, India

Hemant Patel Sumandeep Vidyapeeth, Vadodara, Gujarat, India

Ketki C. Pathak S.C.E.T., Athwalines, Surat, Gujarat, India

Dipti Rana SIESGST, Navi Mumbai, India;
SVNIT, Surat, India

Falguni Ranadive Rishabh Software, Vadodara, Gujarat, India

Kanica Sachdev Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Katra, India

Priti Srinivas Sajja Sardar Patel University, Vallabh Vidyanagar, Anand, India

Oluwafemi A. Sarumi University of Manitoba, Winnipeg, MB, Canada;
The Federal University of Technology—Akure (FUTA), Akure, Nigeria

Jignesh N. Sarvaiya S.V.N.I.T., Ichhanath, Surat, Gujarat, India

Yash Shah Nirma University, Ahmedabad, India

Akil Z. Surti Enlighten Infosystems, Vadodara, Gujarat, India

Pradeep Tiwari Department of Electronics Engineering, Sardar Vallabhbhai Patel National Institute of Technology, Surat, Gujrat, India;
Department of Electronics and Telecommunication Engineering, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, India

Marsha Vegda Nirma University, Ahmedabad, India

Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
ACG	Angle Closure Glaucoma
ACS	American Chemical Society
ADBRF	Adaboost Algorithm with RF
AER	Automatic Emotion (Stress) Recognition
AI	Artificial Intelligence
ALL	Acute Lymphocytic Leukemia
AMD	Age Related Macular Degeneration
AML	Acute Myelogenous Leukemia
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average
ARM	Association Rule Mining
ASR	Automatic Speech Recognition
AUC	Area Under the Curve; Area Under the ROC Curve; Automatic operating Characteristic Curve
BLSTM	Bidirectional LSTM
BMJ	British Medical Journal
BMP	Bitmap
C-EPAC	Coupled Edge Profile Active Contours
CAD	Computer-Aided Design; Computer-Aided Diagnosis
CART	Classification And Regression Tree
CASD	Computer-Aided Skin Disease
CBIS-DDSM	Curated Breast Imaging Subset of DDSM
CET	Central European Time
CHAID	Chi-square Automatic Interaction Detection
CLAMP	Clinical Language Annotation, Modeling, and Processing
CLL	Chronic Lymphocytic Leukemia
CML	Chronic Myelogenous Leukemia

CNN	Computational Neural Network; Convolutional Neural Network
CNVM	Choroidal Neo-Vascular Membrane
Conv.	Convolutional
COVID-19	Coronavirus Disease 2019
Covid-19/COVID-19	Corona Virus Disease found in 2019, linked to family of viruses as Severe Acute Respiratory Syndrome (SARS)
CRF	Conditional Random Field
CRVO	Central Retinal Vein Occlusion
CSR	Central Serous Retinopathy
CSSA	Chaotic Salp Swarm Algorithm
CT scanned images	Computerized Tomography scanned images
CT	Computed Tomography; Computerized Tomography
CUP	Cambridge University Press
CXR	Chest X-ray
CXRs	Chest Radiographs
DCNN	Deep Convolutional Neural Network
DDBJ	DNA Data Bank of Japan
DDSM	Digital Database for Screening Mammography
DEM	Diffused Expectation Maximization
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DNA	Deoxyribonucleic acid
DNN	Deep Neural Network
DOST	Discrete Orthonormal Stock-well Transform
DR	Diabetic Retinopathy
DT	Decision Tree
DTL	Deep Transfer Learning
DTL-CNN	Deep Transfer Learning Convolutional Neural Network
EEG	Electroencephalogram
EMBL-EBI	European Molecular Biology Laboratory—European Bioinformatics Institute
EMBO	European Molecular Biology Organization
EMR	Electronic Medical Record
EST	Expressed Sequence Tag
FAQs	Frequently Asked Questions
FC	Fully Connected Layer
FCBC	Fast Correlation-Based Filter
FCM	Firebase Cloud Messaging
FGLAUC-99	Falguni Glaucoma 99
Fig	Figure
FN	False Negative; False Negative Value
FOA	Fly Optimization Algorithm
FP	False Positive; False Positive Value
FSG	Future Science Group

FT	Fourier Transform
GGO	Ground Glass Opacity
GIP	Gaussian Interaction Profile
GLCM	Gray-Level Co-Occurrence Matrix
GMM	Gaussian Mixture Model
GO	Gene Ontology
GPU	Graphics Processing Unit
GRCh37	Genome Reference Consortium Human Build 37
GRCh38	Genome Reference Consortium Human Build 38
GRCh37.p13	GRCh37 patch 13
GRCh38.p10	GRCh38 patch 10
HB	Hemoglobin
HD	Hausdorff Dimension
HGP	Human Genome Project
HIV	Human Immuno-deficiency Virus
HMI	Human Machine Interfacing
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
HRCT	High-Resolution Computed tomography
HT	Hough Transform
HTK	Hidden Markov Model Tool kKit
Ibk	Instance Based Learner
ICT	Information and Communication Technologies
ID3	Iterative Dichotomiser 3
IDE	Inverse Differentiate Method
IDM	Inverse Difference Moment
IFT	Inverse Fourier Transform
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IOP	Institute of Physics; Intraocular Pressure
iOS	iPhone Operating System
IoT	Internet of Things
ISODATA	Self-organizing Data Analysis Technique
JPEG	Joint Photographic Experts Group; Joint Picture Expert Group
KFCG	Kekre's Fast Codebook Generation
KNN	K Nearest Neighbor
LBG	Linde Buzo Gray
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LDP	Local Directional Pattern
Lib	Library
LOG	Laplacian of Gaussian
LPCC	Linear Prediction Cepstral Coefficients
LR	Logistic Regression
LSTM	Long and Short term Memory

MCH	Mean Corpuscular Hemoglobin
MCS	Multi-Classification
MCV	Mean Corpuscular Volume
MDPI	Multidisciplinary Digital Publishing Institute
ME	Maximum Entropy
MEA	Midpoint Ellipse Algorithm
Mel	Melody
MEMM	Maximum Entropy Hidden Markov Model
MERS	Middle East Respiratory Syndrome
MFCC	Mel-Frequency Cepstral Coefficients
MIAS	Mammographic Image Analysis Society
miRBase	microRNA (miRNA) Database
miRNA	micro-Ribonucleic Acid
ML	Machine Learning
MLP	Multi-Layer Perceptron; Multi-Level Perceptron
MM	Mathematical morphology
Mod	Moderate; it is a fuzzy value of various symptoms such as fever and joint pain
MRI	Magnetic Resonance Imaging
MUC-6	Message Understanding Conferences—6
NB	Naive Bayes
NBML	Naive Bayes-based Machine Learning
NCBI	US National Centre for Biotechnology Information
NCS	Non-Coding Sequence
NEJM	New England Journal of Medicine
NER	Named Entity Recognition
NGS	Next-Generation Sequencing
NIH	National Institutes of Health
NLP	Natural Language Processing
NN	Neural Network
NPDR	Non-Proliferative Diabetic Retinopathy
OAG	Open Angle Glaucoma
OD	Optical Disc
OFR	Open Reading Frame
OOB	Out Of Bag
OS	Operating System
PA	Prophet Algorithm
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PCS	protein-coding sequence
PHOG	Pyramid Histogram of Oriented Gradients
PLOS	Public Library of Science
PMC	PubMed Central is a free digital archive database of full-text scientific literature in biomedical and life sciences
PNAS	Proceedings of the National Academy of Sciences

POAG	Primary Open Angle Glaucoma
PPI	Protein-Protein Interaction
PPV	Positive Predictive Value
PSSM	Position Specific Scoring Matrix
R&D	Research and Development
RAM	Random Access Memory
RBC	Red Blood Cells
RDW	RBC Distribution Width
ReLU	Rectified Linear Unit
ResExLBP	Residual Exemplar Local Binary Pattern
ResNet	Residual Neural Network
RF	Random Forest
RFE	Recursive Feature Elimination
RGB	Red-Green-Blue
RNA	Ribonucleic Acid
ROBC	Region of blood cell
ROC	Receiver Operating Characteristic curve
ROI	Region of Interest
RT-PCR	Real-time Reverse-Transcriptase-Polymerase Chain Reaction
RUP	Rockefeller University Press
SAGE	Sarah and George Publishing
SARS	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SAVEE	Surrey Audio-Visual Expressed Emotion
SCA	Sine Cosine Algorithm
SCCA	Sparse Canonical Correspondence Analysis
SDM	Stimulating discriminant measures
Sec	Seconds
SESSA	Salp Swarm Algorithm
SFTA	Segmentation based Fractal Texture Analysis
SGD	Stochastic Gradient Descent
SMACC	Sequential Maximum Angle Convex Cone
SMO	Sequential Minimal Optimization
SMTT	Self dual Multi-scale Morphological Toggle Block
SNP	Single-Nucleotide Polymorphism
SRGAE	Seed region growing area extraction
SS—SVM	Semi Supervised—SVM
SVM	Support Vector Machine
TB	Tuberculosis
TLA	Transfer learning approach
TN	True Negative; Truly Negative Value
TP	True Positive; Truly Positive Value
UCSC	University of California–Santa Cruz
UI	User Interface

USA	United States of America
V. Low	Very Low; it is a fuzzy value of various symptoms such as fever and joint pain
VAR	Vector Autoregression
VGG-16	Visual Geometry Group Convolutional Neural Network model with 16 Layers depth
VGG-19	Visual Geometry Group Convolutional Neural Network model with 19 Layers depth
VGGNet	Visual Geometry Group Convolutional Neural Network model
VQ	Vector Quantization
WBC	White Blood Cells
WHO	World Health Organization
WP	Wavelet Packet

Chapter 1
Stress Identification from Speech Using Clustering Techniques

Pradeep Tiwari and A. D. Darji

Abstract With the stressful environment of day to day life, pressure in the corporate world and challenges in the educational institutes, more and more children and adults alike are affected by lifestyle diseases. The Identification of the emotional state or stress level of a person has been accepted as an emerging research topic in the domain of Human Machine Interfacing (HMI) as well as psychiatry. The speech has received increased focus as a modality from which reliable information on emotion can be automatically detected. Stress causes variation in the speech produced, which can be measured as negative emotion. If this negative emotion continues for a longer period, it may bring havoc in the life of a person either physically or psychologically. The paper discusses the identification of stress by recognising the emotional state of a person. Herein, four approaches for automatic Emotion Recognition are implemented and their performances such as accuracy and computation time are compared. First approach is Stress/Emotion recognition based on Mel-Frequency Cepstral coefficients (MFCC) feature with Lib-SVM classifier. In other approaches, Vector Quantization (VQ) based clustering technique is used for feature extraction. Three algorithms based on VQ have been explored: (a) Linde-Buzo-Gray (LBG) algorithm, (b) Kekre's Fast Codebook Generation (KFCG) algorithm (c) Modified KFCG. The result obtained indicates that VQ based features perform better in comparison to MFCC, while KFCG modified algorithm gives further better results. The Surrey Audio-Visual Expressed Emotion (SAVEE) database of seven universal emotions and ENTERFACE database with six emotions is used to train and test the multiclass SVM.

P. Tiwari (✉) · A. D. Darji
Department of Electronics Engineering, Sardar Vallabhbhai Patel National Institute of Technology, Surat, Gujrat, India
e-mail: pradeep.tiwari@nmims.edu

A. D. Darji
e-mail: add@eced.svnit.ac.in

P. Tiwari
Department of Electronics and Telecommunication Engineering, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, India

M. Mehta et al. (eds.), *Tracking and Preventing Diseases with Artificial Intelligence*, Intelligent Systems Reference Library 206,
https://doi.org/10.1007/978-3-030-76732-7_1

1.1 Introduction

Mental stress is a serious problem nowadays that not only affects the capacity, performance and mood of an individual, but also induces physical and mental health issues [1]. Under several stressed circumstances or emotions, the attributes of speech signals vary [2]. Stressed speech is characterized as the speech generated under any situation that leads the speaker to vary the speech from the neutral condition in the production of speech [3]. If a speech generated is in a 'quiet place' with no work duties, therefore the generated speech is assumed to be neutral. Stress can be categorized as (a) Emotionally driven stress: Speech generated by a shift in the speaker's mental or psychological condition like angry speech, happy speech, etc. (b) External stress triggered by the atmosphere such as Lombard speech (c) Pathological stressed speech such as Cold influenced speech, Senior Citizens Speech. In this work, emotionally driven stressed speech and External stress triggered by the atmosphere are considered. Stress unlike physiological diseases does not show symptoms at an early stage, so it can be cured before it is a massacre. Stress can be identified with the help of seven universal human emotions like fear, anger, disgust, happiness, contempt, and surprise and sadness as suggested by Ekman and Friesen [4]. Speech being non-invasive, non-intrusive in nature attracts the majority of the researchers and deals with identification of stress or emotion. For the applications like Human Machine Interface to work on low-cost processors or mobile applications, it becomes challenging to obtain the accurate results in real time. Thus, this paper focuses on the investigation of feature extraction techniques which increases the recognition accuracy and decreases the computational time by suitable modification in features like clustering, thus arriving at a simpler approach to perform real time fast and efficient emotion classification. Clustering reduces the size of training vector by quantizing it into clusters. Hence, the major contribution of this paper is it investigates the techniques which reduce the complexity involved while training and testing of a classification model considerably which further decreases the computation time without compromising with the accuracy. Four approaches for automatic Emotion Recognition are implemented in this paper and their performances such as accuracy and computation time are compared. First approach is Stress/Emotion recognition based on Mel-Frequency Cepstral coefficients (MFCC) feature with Lib-SVM classifier. In other approaches, Vector Quantization (VQ) based clustering technique is used for feature extraction. Three algorithms based on VQ have been explored: (a) Linde-Buzo-Gray (LBG) algorithm, (b) Kekre's Fast Codebook Generation (KFCG) algorithm (c) Modified KFCG. The result obtained shows that the performance of VQ based features is better in comparison to MFCC, while KFCG modified algorithm shows further improvement in the results. Results also illustrates that the clustering technique reduces the possible overfitting, bias and variance issues along with reducing the dimensionality of the features thus improving the results.

The remaining paper is organized as follows.Related work is discussed in Sect. 1.2. Section 1.3 details the experiments conducted while the results and its analysis is discussed in Sect. 1.4. Section 1.5 describes the future scope and concludes this work.

1.2 Related Work

Lots of researchers have contributed to emotion and stress identification area in past one decade. Ramamohan et al. [4] have utilized Sinusoidal features which can be characterised by its Amplitude, Frequency and Phase as features. Its accuracy is calculated for four emotions such as Anger, Neutral, Happiness and Compassion with Vector Quantisation (VQ) and Hidden Markov model (HMM) classification algorithm which shows better results compared to cepstral features and the linear prediction algorithm. Shukla et al. [5] considered a database consisting of five emotions, namely angry, neutral, happy, sad and Lombard, using 33 words. VQ and Hidden Markov model (HMM) were used as classification models for 13 dimensional MFCC features. The result obtained was 54.65% for VQ and 56.02% for HMM, while a result of 59.44% was found to give human classification of stress. According to the survey conducted by Hasrul et al. [6] for emotion recognition with prosodic features such as pitch, MFCCs with Gaussian Mixture Model (GMM), formants with SVM and energy with GMM, energy with GMM gave the best result of 92.3%. Shilpi et al. in 2015 [7] proved that speech signals combined with textual information improves the accuracy of emotion recognition. MFCC and Mel-Energy Spectrum Dynamic Coefficients features with Lib-SVM classifier was used by Chavan et al. [8] for happiness, sad, anger, neutral, fear, from Berlin database with 93.75% accuracy. A comparative study for word level and sentence level utterances from the SUSE database was carried out by Sudhakar et al. in 2014 [9]. Linear Prediction Cepstral Coefficients (LPCC) and MFCC features were extracted. They conclude that word utterances performed better than sentences and 2nd, 3rd and 4th order coefficients also gave comparable results to 12/13 order coefficients. A novel technique was proposed by Amiya et al. in 2015 [10]. They combined prosody features, quality features, derived features and dynamic feature for robust emotion recognition. Anger, disgust, fear, happy, neutral, sad and surprise emotions were classified using SVM. Revathy et al. in 2015 [11] discusses the effectiveness of Hidden Markov Model tool kit (HTK) for speaker independent emotion recognition systems for EMO-DB database with 68% accuracy. A novel WP-based feature called Discriminative band WP power coefficients was introduced by Yongming et al. in 2015 [12] for emotion recognition. These features gave improved performance over MFCC. El Ayadi et al. [13] explains the features, formants, and energy-dependent properties related to pitch contribute to the recognition of speech emotion. For the SAVEE Database, Sanaul et al. proposed the speaker-dependent feature by case recommending feature selection on 106-dimensional audio features [14]. In addition, Davood et al. used Fast Correlation-Based Filter (FCBF) feature selection on MFCC, Formants, and related statistical features on the SAVEE database with an average accuracy of 53% for fuzzy ARTMAP neural networks in 2017 [15]. Significant changes were observed over spectral features when weighted MFCC features were combined with spectral and prosody features [16]. Deb et al. [17] suggested region flipping based classification strategy using vowel-like regions and non-vowel-like regions using the Extreme Learning Machine classification model on the EMO-DB database. Wissam et. al. build the SVM model by merging neurogram features and traditional speech features [18].

This work is targetted for low-cost processors or mobile applications where it becomes challenging to obtain the accurate results in real time. Thus, there was a need to investigate feature extraction techniques which increases the recognition accuracy and decreases the computational time. In the above mentioned literature, various approaches were considered but the computation time requirement in those approaches are more. Since, clustering reduces the size of training vector by quantizing it into clusters. Hence,this paper investigates the techniques which reduce the complexity involved while training and testing of a classification model considerably which further decreases the computation time without compromising with the accuracy. Experimental results show that the proposed feature considerably improves emotion recognition performance over MFCC feature.

1.3 Stress Identification System Setup

The Block diagram of stress identification system set-up is represented by Fig. 1.1.

The first step in stress identification system is speech signal acquisition which is obtained from two standard databases: (i) The Surrey Audio-Visual Expressed Emotion (SAVEE) database of seven universal emotion (neutral, fear, disgust, happy, anger, sad and surprise) (ii) ENTERFACE database with 6 emotion (neutral, fear, disgust, happy, anger, sad and surprise).

The acquired signal will have lots of unwanted part like silence, surrounding noise, dc offset values etc., and thus it is required to pre-process the speech signal. Pre-processing includes three steps: (a) Eliminating the redundant information in the signal (b) Removal of dc offset values which does not carry any information by the process called normalisation (c) Pre-emphasizing the speech signal by using a high pass filter since the speech produced is deemphasized at glottis. The next step is extracting the feature from speech signal. There are various features extraction techniques like Cepstral Co-efficients, MFCC and LPC coefficients which can be

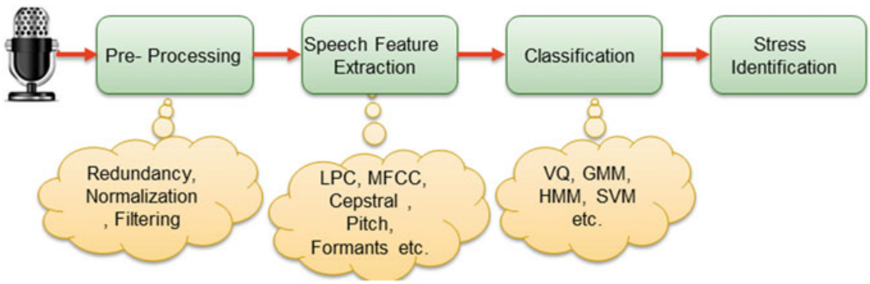

Fig. 1.1 Block diagram of stress identification system

applied to get feature vectors. There are two types of speech features which have been used by the researchers, (a) Prosodic speech features such as pitch and energy, also called local features. (b) Statistic or transform based features such as MFCC, Wavelet, also known as global features. MFCC in Fig. 1.2 and Vector Quantisation based features are considered in this research. Further, a pattern classifier called support vector machines (SVM) decides the emotion class of the utterance.

1.3.1 Signal Aquisition and Pre-processings

The first step is speech signal acquisition which is accomplished using standard database. The speech signal which is employed for AER is from standard SAVEE database of seven universal emotions (Anger, disgust, fear, happy, neutral, sad and surprise). The acquired signal would consist of unwanted part like silence, surrounding noise and dc offset values provided by microphone while recording, so it is required to pre-process the speech signal. The second step is extracting the feature from speech signal, wherein algorithms of various features extraction techniques like Mel frequency cepstral co-efficients from speech and facial landmarks from image are utilised to get feature vectors. These feature vectors will be used in third step where classifier models like Support vector machine (SVM) would classify the different emotion classes. The Pre-processing includes normalization and pre-emphasis.

1.3.2 Speech Feature Extraction

The second step is extracting the feature from speech signal, wherein algorithms which can provide intra-class resemblance and inter-class discrimination are applied to get feature vectors. The performance of any stress/ emotion identification system mainly depends on features extracted from speech emotion signal. Mel-frequency cepstral co-efficients (MFCC) is extracted for ASR and AER.

1.3.2.1 Mel-Frequency Cepstral Coefficients

This is a most widely used speech feature extraction technique found with the multiplication of Mel-filter bank with the frequency attribute of the signal called Power spectrum [7]. MFCC is based on human hearing perceptions, i.e. MFCC is calculated by considering the variation of the human ear's critical bandwidth with respect to frequency. The MFCC feature extraction technique is as shown in Fig. 1.2.

Mel-Scale or Melody-scale is computed if frequency f is given in Hz, with Eq. 1.1 is used.

$$Sk = Mel(f) = 2595 * (log_{10}(1 + \frac{f}{den})) \tag{1.1}$$

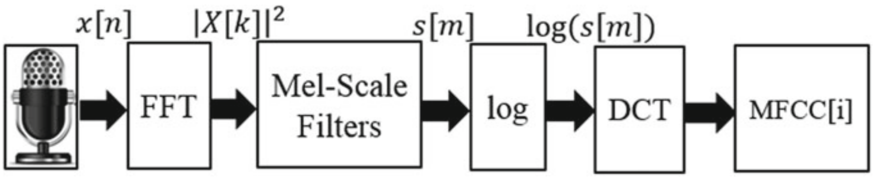

Fig. 1.2 MFCC Feature Extraction

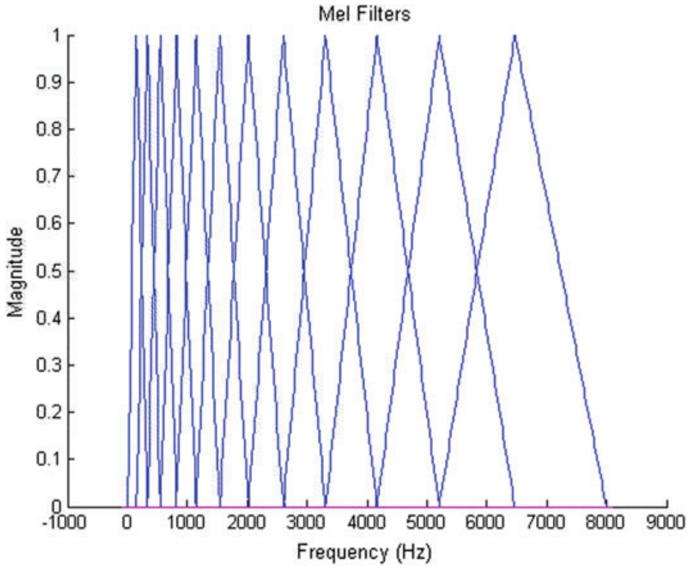

Fig. 1.3 Mel Filter Bank

The Mel filter bank obtained is given in Fig. 1.3.

First, the logarithm of the absolute value of the fast Fourier transform of input signal $x[n]$ is calculated whose inverse fast Fourier transform gives Cepstrum as shown in Eq. 1.2.

$$Cepstrum = IFT[abs(log(FT(x[n])))] \tag{1.2}$$

where, $FT(x[n])$ indicates to the fast Fourier transform of speech signal and $IFT(signal)$ means the inverse fast Fourier transform of the speech signal. The short time Fourier transform for a frame is illustrated in Eq. 1.3.

$$X_a[k] = \sum_{n=0}^{N-1} x[n]e^{\frac{-j2\pi nk}{N}}, \quad 0 \le k \le N \tag{1.3}$$

As $X_a[k]^2$ is known as Power spectrum and if it is applied to Mel frequency filter bank $H_m[k]$ consist of triangular filters, it results into Mel-frequency power spectrum as provided in Eq. 1.4.

$$S[n] = \sum_{n=0}^{N-1} X_a[k]^2 H_m[k], \quad 0 \leq m \leq M \tag{1.4}$$

Now, the log Mel-frequency power spectrum output is returned back to time domain by utilizing a compression algorithm called discrete cosine transform on $S[m]$. This gives MFCC calculated as shown in Eq. 1.5.

$$MFCC[i] = \sum_{m=1}^{M} log(S[m]) \cos[i(m - \frac{1}{2})\frac{\pi}{M}] \quad i = 1, 2, \ldots, L \tag{1.5}$$

The value of L is 13 i.e. it produces 13 MFCC coefficients for each frame and M indicates the length of the speech frames. The diagram shown in Fig. 1.4 represents the input speech signal and output as extracted MFCC features.

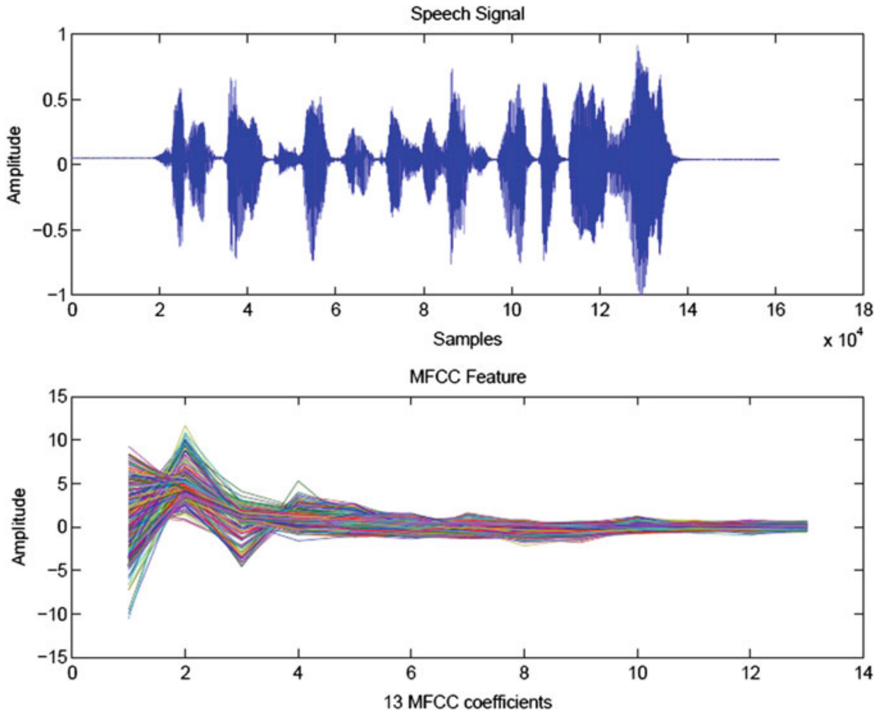

Fig. 1.4 MFCC Feature Extraction

1.3.2.2 VQ Based Features

Vector Quantization [19] reduces the size of training vector by quantizing it into clusters called codebook as shown in Fig. 1.5. It reduces the complexity involved while training and testing considerably as explained in Fig. 1.6.

Since this technique reduces the size of training vectors by quantizing it, it will be applied after extraction of MFCC feature. The complexity involved while training

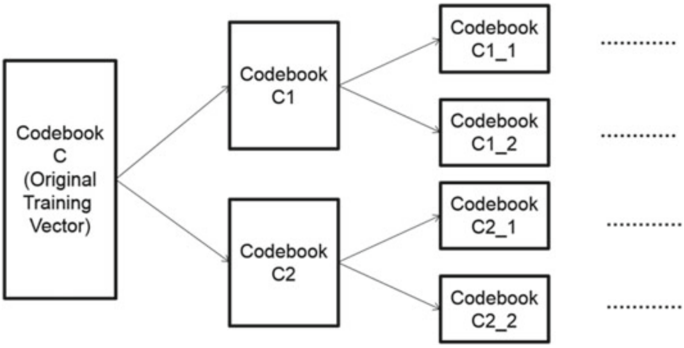

Fig. 1.5 Block diagram of stress identification system

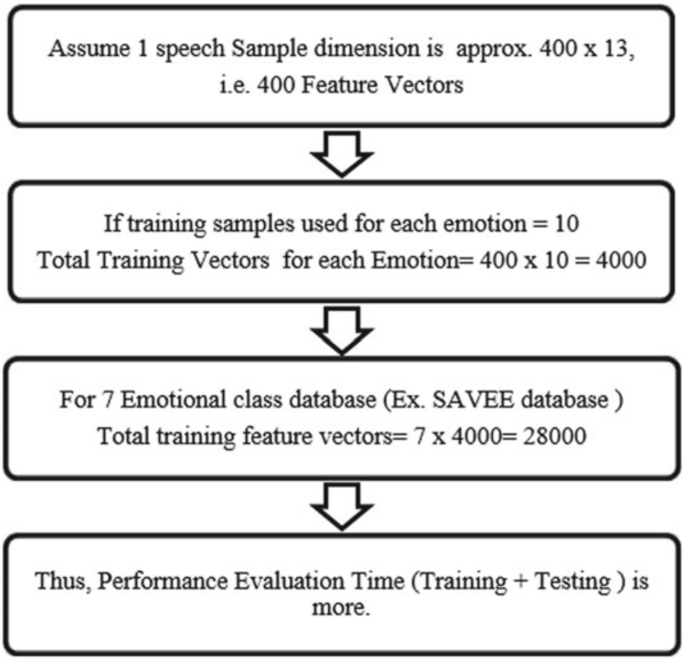

Fig. 1.6 Need of training data vector size reduction