Studies in Big Data 85

Kalaiselvi Geetha Manoharan
Jawaharlal Arun Nehru
Sivaraman Balasubramanian *Editors*

# Artificial Intelligence and IoT

## Smart Convergence for Eco-friendly Topography

Springer

# Studies in Big Data

Volume 85

The series "Studies in Big Data" (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are reviewed in a single blind peer review process.

Indexed by zbMATH.

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/11970

Kalaiselvi Geetha Manoharan ·
Jawaharlal Arun Nehru ·
Sivaraman Balasubramanian
Editors

# Artificial Intelligence and IoT

## Smart Convergence for Eco-friendly Topography

Springer

*Editors*
Kalaiselvi Geetha Manoharan
Department of Computer Science and Engineering
Annamalai University
Chidambaram, Tamil Nadu, India

Jawaharlal Arun Nehru
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, Tamil Nadu, India

Sivaraman Balasubramanian
Department of Mechanical Engineering
Annamalai University
Chidambaram, Tamil Nadu, India

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-981-33-6399-1 ISBN 978-981-33-6400-4 (eBook)
https://doi.org/10.1007/978-981-33-6400-4



This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

Internet of Things (IoT) is getting smarter with artificial intelligence (AI). Companies are incorporating artificial intelligence into their IoT applications and finding their ability grows, which includes improved operational efficiency. Remarkably, these two technology concepts absolutely complement each other. As the number of connected devices grows, never-ending information surge is streamed from IoT gadgets and sensors, empowering the globe in all aspects. IoT puts the 'smart' in homes, facilitates products impress customers, creates industrial equipment safer, gives real-time updates about patients' health, and a lot more. AI is an efficient elucidation to manage and organize numerous connected IoT components. On top of that, unrestricted processing and learning capabilities of AI are vital for making sense of heaps of data broadcasted by IoT appliances. As everything is connected to the Internet, the proposal of a future connected world is becoming a practicality. Of course in reality, a lot of challenges still remain.

The objective of this edited book is to provide an insight to the targeted readers from varying disciplines who are interested in implementing the smart planet/environments via AI-empowered IoT. To achieve these objectives, the book provides an imminent into IoT intelligence to address the foremost challenges in realizing the fusion of AI in IoT-based applications together with challenges that diverge from cost and energy competence to accessibility to service eminence in multidisciplinary manner. To stay abreast with the progress in a consistent style, we strove to keep the book reader friendly so that it can serve as a handbook for many researchers. This has been done to make the edited book more flexible and to stimulate further research interest in topics. We trust, our effort can make this collection interesting and highly attract the student pursuing pre-research, research and even master. We trust and hope that this edited book will help the researchers, who have interest in fusing AI with IoT and its applications to keep insight into recent advances in real-life applications.

Emotion recognition system using deep learning networks is discussed in Chap. 1, "Smart IOT Multimodal Emotion Recognition System Using Deep Learning Networks." Smart IoT-based devices are used to extract emotions and sentiment metrics of members from e-learning content. ELMs are utilized to combine the output obtained from the CNN and RNN. Multimodal information like text, image, and acoustic data with respect to time is extracted by CNN, and the extracted output is processed by both CNN and RNN for sentiment analysis. Ensemble-based algorithms are used to train the CNN. Ensemble algorithms are used in the hidden neurons of CNN to carry the time series information to the RNN with the help of stumps and identity activation function. The parameters of models are computed by minimizing not only the classification error (reconstruction error in the case of the ELM-CNN) but also the network weights which lead to better generalization performance.

The lower sensitive boundary pixels makes it difficult to identify the tumor areas in the Glioma brain image. Internet of Things-based system is proposed to identify the brain tumor in Chap. 2 “Internet of Things-based Healthcare System for Glioma Brain Tumor Detection and Classification Model Hierarchical Convolution Neural Network.” A novel Glioma brain tumor (GBT) recognition and classification model by means of hierarchical convolution neural network (HCNN) is presented. Nonsub sampled contourlet transform (NSCT) is utilized for enhancing the GBT images, and texture features are extracted for analyzing the brain images. Hierarchical convolution neural network is used to classify the images after segmentation using the BRATS dataset. The experimental results show the good performance of the model.

There is a strong interest to use social networking to enhance the communications among different IoT. Health monitoring system is explained in Chap. 3 “Artificial Intelligence and IoT Framework for the Health Monitoring System.” Recently, Internet of Everything (IoE) and Internet of Nano-Technology (IoNT) are the most prominently distributed knowledge seen. IoNT is widely seen in medical smart watch calorimeter and pulse rate instrumentation. Strong interest is seen to use social networking to enhance the communication using IoT using a new paradigm named Social Internet of Things (SIoT). For health monitoring, SIoT systems enable AI doctor to collect data over social networks and automate the process of transferring data over social networks for disease diagnosis. An adaptive trust management protocol is developed for SIoT systems toward edge service management.

With the advent of modern technology and world wide web, data is being collected from everywhere. This collected data is huge and is temporal, spatial, structured, semi-structured, un-structured, and multi-structured. For proper usage of this data, classification is essential for knowledge discovery and intelligent decision making. Chapter 4 “Big Data Classification: Applications and Challenges,” provides an overview of the prominent areas of big data and the challenges in handling the big data such as storing, searching, visualization, security, privacy. Further, it explains the different techniques for big data classification. Various real-world applications like big data classification, including health care, user data analysis, network traffic analysis, etc., are elaborately reviewed and discussed. The problems with imbalanced classification and the techniques for handling are also discussed.

The real difficulty of artificial intelligence continues to prove by doing the tasks which are challenging to describe formally. Deep learning algorithms are related to how the nervous system structured/ works where each neuron connected each other and passing information. Convolutional neural networks are frequently used to analyse image representation. A survey of the deep learning architecture is discussed in Chap. 5, "A Survey on Recent Deep Learning Architectures." The chapter discusses the fundamental building blocks of the convolutional neural network architecture. In addition, it explains the various architectures including LeNet, AlexNet, and so on. The chapter also compares these architectures based on the learnable parameters.

Chapter 6 "Research Progress of MapReduce Big Data Processing Platform and Algorithm" reviews the research progress of big data processing platforms and algorithms based on MapReduce programming model in recent years. MapReduce-based big data processing platform is introduced, analyzed, and the chapter compared their implementation principles and applicable scenarios, based on MapReduce big data analysis algorithms, including search algorithms, data cleaning/transformation algorithms, aggregation algorithms, join algorithms, sorting algorithms, preference queries, optimization calculations. In MapReduce implementation, the factors affecting the performance of the algorithm are analysed. The processing algorithm is abstracted as an external memory algorithm, and the characteristics of the external storage algorithm are sorted out. The research ideas and problems of the performance optimization method of the universal external memory algorithm are proposed. The existing large data processing platform and algorithm research mostly focus on platform dynamic performance optimization based on resource allocation and task scheduling, specific algorithm parallelization. The proposed performance optimization of the external memory algorithm is a static optimization method, which is a good complement to the existing research, providing researchers with a broad research study space.

Predicting the intensity of cyclones is vital for safeguarding the coastal regions. A devastation effect is produced by the tropical cyclone (TC) and is a constant threat to the people in and around coastal area. In India, nearly 10% of the world's tropical cyclones have been experienced, wherein the major portion is affected in the eastern side of India which is in the Bay of Bengal. Every year, India faces 5–6 cyclones and out of which 2–3 are highly destructive. Ninety percent of non-cyclones do not get into the sea because of cold waters, heavy wind shear, or if they come in close proximity to the equator. Accurate prediction and estimation of TC intensity will save many lives and property. TC genesis is a circuitous task which constitutes inputs from various sources at different levels. Accurate prediction and estimation of TC intensity will save many lives and property. Chapter 7 "Recognising Tropical Cyclone Formation from Satellite Image Data" proposes an approach for predicting the tropical cyclone along the coastal area. Classification of the various intensity levels of clouds' structure is done. The cylindrical symmetry about the axis is calibrated to understand the cloud formation in the primeval phase of the development process wherein the intensity scale of TCs is predicted. The axisymmetry of a cloud cluster is quantified using the numerical study of the inclination of antenna temperature from the satellite to the ground. The acclivity of the antenna temperatures from the GOES satellite in the IR images is calculated from the horizontal and vertical asymptotes by using Sobel edge detection method, and then the deflections of each and every acclivity is being calculated from the benchmark. The squared deviations of these angles to measure the symmetry about an axis, acts as the end result is estimated. From the metric of the axisymmetric structure and the wind speed, the wildness of the TC is classified using SVM and neural network.

In recent years, it is mandatory to have automatic road signal recognition system that helps in designing automated vehicles. The objective of this Chap. 8 entitled "A CNN-based Traffic Sign Recognition System" is to detect and recognize the traffic signs using two methods. Traffic signs are silent speakers to the road users. It gives advance information about road conditions ahead. In this chapter, an effective method is proposed to identify the traffic signs in an automated manner. The traffic sign recognition for advanced driver-assistance system of Method I comprises of three stages: (1) detection using Haar-like features with AdaBoost method, (2) feature extraction using speeded up robust features, and (3) classification using support vector machine. The second approach comprises of two stages: (1) detection of traffic sign using convolution neural network and (2) classification using convolution neural network. The performance of the system is analyzed. From the experimental results, it is observed that Method II using convolution neural network gives a better performance compared with Method I.

Smart Cities is venturing into reality as the center shifts from hypothesis to real advancement. Sustainable power sources wind vitality, solar-powered vitality, hydro-vitality, and warm vitality are the lights of future on the grounds that the customary sources like fossil, coal, and oil fills are constrained and are arriving at an end as a result of the expanding request. So as to fabricate savvy medical clinics, structures, plants, traffic and transportations, a dependable strong, effective and smooth vitality stream is utilized. All these computerized administrations are foreseen to run without disturbances by the utilization of electrical control frameworks and keen vitality, which are viewed as the spines of such urban areas. So as to keep the shrewd city administrations interconnected and matched up, IoT assumes a key job. Through the span of this piece, we will go over the job of IoT and sustainable power sources. The usage of sustainable power source would not be so fruitful without IoT. This review Chap. 9 "Internet of Things (IoT)-based Renewable Energy and Sustainable Power Sources," discusses the correlation between the a portion of the sustainable power sources, for example, solar oriented, wind, hydro, and warm vitality with Internet of Things (IoT). Utilizing the IoT, a framework will be less defenseless to blackouts and efficiency issues (coming about because of personal time) and possibly exorbitant security breaks. By introducing a basic, start to finish arrangement utilizing IoT, one can deal with the biggest solar-based frameworks on the planet, even with a large number of individual gadgets associated with the system.

The well-managed and maintenance structures are the safest structures among all the structures such as roads, buildings, tunnels, bridges, and dams in civil engineering. Artificial intelligence (AI) has been used widely in all sectors of education, science, engineering, and industries. AI is an alternative approach to the classical modeling technique (MT). AI technique is developed by the department of computer science which works on the theory of human brain ability. As compared to traditional methods, the AI approach provides a solution to many complex civil engineering problems, especially in the area of concrete technology, structural health monitoring, geotechnical engineering, transportation engineering, and structural analysis. In addition to this, AI approach also works on those areas where physical testing is not possible, which ultimately reduces human efforts, time, and cost. The core advantage of using AI technique is that it reduces the errors and provides quick results by quick decision making (QDM) power. In the field of civil engineering, some of the most frequently used artificial intelligence techniques are machine learning (ML), deep learning (DL), and pattern recognition (PR). Chapter 10 "An Approach-Driven: Use of Artificial Intelligence and Its Applications in Civil Engineering" provides a detailed overview of use of these AI techniques in different aspects/field of civil engineering starting from the general introduction of AI, and the chapter ends up with the inclusion of all possible areas where the use of AI techniques has been widely accepted by the construction team in smart cities to achieve safe, sustainable, precise, and accurate results.

Deep learning is the next phase of machine learning in recent decades. It has significantly changed the way in which computer systems interpret human-centric content such as images, video, speech, and audio. Different models have been introduced based on learning techniques such as supervised, unsupervised, reinforcement, and it is expected to accelerate and create even more innovative models in the coming years. With the rise of the Internet of Things (IoT), many real-time applications collect data about people and their environment using IoT sensors and feed them into deep learning models to enhance the intelligence and the capabilities of an application, for offering better recommendations and service. Experimental results of most of these applications look promising when compared with traditional machine learning approaches. Chapter 11, "Deep Learning Algorithm and Its Applications to IoT and Computer Vision" has provided a general review of deep learning, various frameworks of deep learning, applications, and how the deep learning is related with IoT over past years. So, the main objective of this chapter is to make a self-contained review of deep learning (DL) models, starting with the convolutional neural network (CNN), recurrent neural network (RNN), long-term short memory (LSTM), autoencoder (AE), generative adversarial network (GAN), and deep reinforcement learning (DRL). Additionally, for providing better understanding of the models and their efficiency, we have added recently developed DL tools or framework and their applications.

Chapter 12 entitled "Exploiting Internet of Things (IOT) Services to Emotion Recognition System—An Application for Autism Children" discusses an approach for recognizing the emotions of autism children automatically. With the advanced technological innovations, Internet of Things (IoT) is taking an important part in every domain. The aim of the proposed work is to monitor the mental status of patients with autism. Autism spectrum disorder (ASD) is a chronic neurodevelopmental sickness that includes core deficiencies in interpersonal verbal exchange, social interactions, minimal, repetitive mannerisms, and interests. Symptoms of social communication deficits and restrictive, repetitive patterns of behavior emerge early in the development of a child. A patient's doctor or family members face difficulties in interpreting autism adult behaviour, since the emotions of autism children are to be understood only through their repetitive actions. Recently, deep learning algorithms have become popular across hierarchical architectures for human emotion recognition to analyze physio-logical signals in multiple modalities. Deep learning, however, is mainly effective for the extraction of deep features. In this chapter, ER16Net and AlexNet models are used with different parameters for emotion recognition from body movements of autism children. The recorded autism children emotion videos are used for this experiment. The experiment is evaluated with two deep learning models. The performance measures are calculated using qualitative and quantitative analysis.

Chidambaram, India — Kalaiselvi Geetha Manoharan
Chennai, India — Jawaharlal Arun Nehru
Chidambaram, India — Sivaraman Balasubramanian

# Acknowledgements

It is with great sense of satisfaction that we present our edited book and wish to express our views to all those who helped us both direct and indirect ways to complete this project. First and foremost, we praise and wholeheartedly thank the Almighty God, which has been unfailing source of strength, comfort, and inspiration in the completion of this edited book. We also immensely acknowledge the authorities of Annamalai University for permitting us to take up this project.

While writing, contributors have referred to several books and journals, and we take this opportunity to thank all those authors and publishers. We are extremely thankful to the reviewers for their constant support during the process of evaluation. Special mention should be made of the timely help given by different persons during the project work, those whose names are not mentioned here. Last but not least, we thank the series editor "Janusz Kacprzyk" and the production team of Springer for encouraging us and extending their cooperation and help for a timely completion of this edited book. We trust and hope that it will be appreciated by many readers.

# Contents

# About the Editors

**Dr. Kalaiselvi Geetha Manoharan** is Professor with 26 years of experience demonstrating consistent success as an educator at the undergraduate and postgraduate education in Computer Science and Engineering, Annamalai University. She has outstanding track record in assuring student success. She was committed to conceiving and building programs from the ground up through proven competencies in board of studies, grant writing, administration, program management, staff development and empowerment. She has strong attitude toward teaching and in motivating students to develop their expertise in major areas like AI, machine learning, computer vision, cloud computing and IoT. Her active participation is in continued learning through conferences and professional research. She believes in students' abilities to learn and an inherent thirst for knowledge with the right environment. She actively works to connect students to their material to transform lives. She is associated with CSI and ISTE professional bodies.

**Dr. Jawaharlal Arun Nehru** is currently working as Assistant Professor in SRM Institute of Science and Technology, with 4 years of professional teaching experience in the Department of Computer Science and Engineering, and committed to comprehensive problem-solving abilities, ability to cope and work with people in groups, leadership quality and analytical and interpersonal skills. He has strong attitude toward teaching and in motivating students to develop their expertise in major areas like deep learning, machine learning, AWS cloud computing and IoT. His active participation is in continuous learning through workshops, seminars, conferences and professional research and associated with ICTACT, IET, IAENG, IEDRC and ISTE professional bodies.

**Dr. Sivaraman Balasubramanian** is Professor in the Department of Mechanical Engineering, Annamalai University with 33 years of teaching experience at both the undergraduate and postgraduate levels. He always strives hard to maintain outstanding track record in assuring student success and progression. He has commitment to conceive and build programs from the scratch. He contributed a lot in administration, program management, staff development and empowerment. He has got positive approach and attitude in teaching students to develop their knowledge in major areas like CFD, heat transfer, renewable energy and data analysis. He actively participates in continued learning through conferences and professional research. He believes in students' abilities to learn and an inherent thirst for knowledge with the right environment. He actively works to connect students to their material to transform lives. He is associated with ISTE professional body.

# Part I
# Artificial Intelligence (AI) and Internet of Things (IoT) Framework

# Chapter 1
# Smart IoT Multimodal Emotion Recognition System Using Deep Learning Networks

**V. J. Aiswaryadevi, G. Priyanka, S. Sathya Bama, S. Kiruthika, S. Soundarya, M. Sruthi, and N. Nataraj**

**Abstract** Smart IOT-based devices are used to extract the emotions and sentiment metrics expressed by the course members for calculating the utility function value for each e-learning content. This is mainly used for rating the E-learning content available on online sources. ELMs are used to combine the output of classifier models implemented by CNN and RNN. Text, image, and acoustic data with respect to time are extracted by CNN, and the extracted output is processed by both CNN and RNN for sentiment analysis. The hidden layers of CNN are trained with ensemble-based algorithms, namely EN-ELM and progressive EN-ELM. The output of CNN and RNN is fed to the two consecutive ELM further whose output is fed to the SVM classifier. Unlike other CNN algorithms, ensemble algorithms are used in the hidden neurons of CNN to carry the time series information to the RNN with the help of stumps and identity activation function. The accuracy of progressive convolution neural network is compared to the accuracy of CNN with RNN. Arousal in the accuracy level is detected with 93–97%.

**Keywords** ELM · CNN · RNN · Progressive EN-ELM

V. J. Aiswaryadevi (✉)
Department of Information Technology, Dr NGP Institute of Technology, Coimbatore 641 035, India
e-mail: aiswarya.devi@live.com

N. Nataraj
Department of Information Technology, Bannari Amman Institute of Technology, Erode, India
e-mail: n.nataraj08@gmail.com

G. Priyanka · S. Sathya Bama · S. Kiruthika · S. Soundarya · M. Sruthi
Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India



K. G. Manoharan et al. (eds.), *Artificial Intelligence and IoT*, Studies in Big Data 85,
https://doi.org/10.1007/978-981-33-6400-4_1

## 1.1 Introduction

Emotions are the way of expressing the human feelings. There will not be any differences between the human beings and the non-living things, if they fail to express their emotions. Emotions are the mixture of happiness, sadness, anger, disgust, fear, surprise, and so on. Emotions should be in a neutral state for a perfect well-being. Human life progressions are purely depends on emotions. Emotions can be identified with help of facial muscle movements. Emotions are not only identified based on facial expressions but also, it can be detected by observing a person's activity on social media, emoji's that he/she is using during their virtual conversation, body language, gestures and voice tone during a phone call. Emotion recognition plays a vital role in several industries such as health care, advertisements, e-commerce, entertainment. Hence, all these data can be captured and analyzed using IoT and deep learning methods.

Emotion detection is done using IOT sensing functions such as face detection, emotion recognition, age prediction, gender prediction and so on. The state-of-the-art CMOS image sensors include highly flexible pixel designs that can logically sense instead of capturing the imaging data, while being coupled with brilliant vision processors. Omron started its human vision component (HVC) module way back in 2013. It has been proceeding since then through better components as well as improving upon its underlying OKAO face recognition algorithm. It has been developed to the point where different facial points can be tracked for interpreting micro-expressions and eye movements. It is also capable of recognizing human emotions, moods and even intentions. Poria et al. [1] express the usage of RNN and ReLu step function for autoencoders, and each acoustic frequency bands are equally spaced on the mel scale using MFCC. It assumes that every coefficient dimension will have zero mean and unit variance. Two consecutive ELMs are used for extracting the accuracy in modalities of information extracted by MFCC, CNN and OpenCV commands. Ensemble algorithms EN-ELM and progressive EN-ELM are used for training the hidden layers of CNN. The HMM models are used for training the NLP models of acoustic frames with the emission and transition probabilities. Hori et al. [2] depict the adaptive HMM models extracting both image and acoustic transition in data retrieved.

Section 1.2.1 covers the video input processing into image frames with face recognition followed by Sect. 4 with inferences about the results of CNN fed into RNN and Sect. 1.2.2 illustrating the usage of ELMs in the network. The acoustic frames extracted from the MFCC spectrum is taken into CNN for emotion recognition which is shown in Sect. 1.3 with feature scaling spectrogram. Ensemble-based classification is used for multimodal sentiment recognition most commonly. Section 1.3.3 deals with the results and discussion obtained.

## 1.2 Problem Formulation

Advanced driver-assistance system (ADAS) which is an inward-looking ADAS camera can avert distracted driving or sleepiness by identifying signs of fatigue or inattentiveness on a driver's face. Law enforcement officers can extract hints from facial expressions, illumination patterns, and blink and gaze estimation to predict the purpose. In business, organizations could use facial recognition data to enhance sales calls and business conciliations. In case when image sensors are reaching the boundaries of pixel size reduction, there is a chance they are moving past the conventional Bayer filter patterns. As an outcome, they can include lucid pixels and thus successfully function under low-light conditions. That is to say, image sensors are turning out to be smarter. The imaging results that include computational photography and deep learning algorithms can assist doctors take on facial recognition to better handle with anger management, autism and depression. OmniVision's OV2744 1.4-micron PureCel image sensor enables biometric capabilities by providing both a high-quality infrared as well as 1080p high-definition RGB output at 60 fps.

Gaze information, blink estimation, face direction estimation and expression estimation together constitutes the emotion or sentiment analysis.

The polarities analyzed are categorized into weak student or negative, good or positive and neutral in listening to the E-content. Nowadays, E-course learners are more prominently seen throughout the universe, where the course enrollers are yet

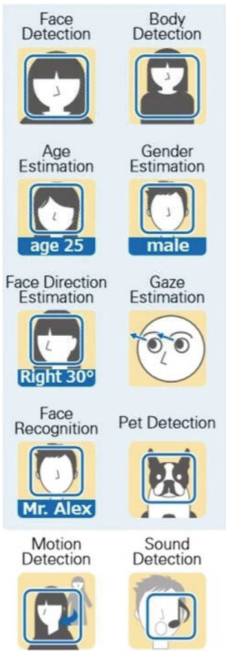


**Fig. 1.1** IOT sensors for emotion identification

to be identified with their unique characteristics in learning. If the course learners are weak, they need slow lectures and fast learners will be feeling better with the regular normal speed of the lecture and in contrast slow learners will be in need of some further reference links to the contents where they feel that they have least understood. Utility function is evaluated at each layer of feature extraction in CNN max pooling. Several layers of LSTM are interpreted in the bidirectional mode with the help of bidirectional LSTM in CNN for feature extraction in images (Fig. 1.1).

### *1.2.1 Emotion Recognition Tools and Algorithm*

Visual data is distributed into spatial dimension of standardized pixels, and the face data is alone filtered from the spatial dimension. The main contribution of the emotion recognition algorithm is to identify the emotion class to which the face lies under. The fusion of min-max algorithm and nearest neighbor algorithm [3] is used to classify the emotions expressed by the face. The images with 256 $\times$ 256 pixels are pruned into a size of 101 $\times$ 114 pixels maintaining only the appropriate information on the face area. Normalization is important to diminish inter-class feature discrimination that can be observed as intensity offsets. Gaussian normalization is applied using the changes in openCV program, whereas intensity offsets are unvaried within the local region. The input image is depicted as $x(i,j)$ and $y(i,j)$ is the normalized output image. The openCV Python code takes care of the image frames extraction from video data [4, 5]. Duong et al. [6] derived the spatial features with respect to the correlation coefficients of Gaussian parameters [7] with CNN gated recurrent neural networks.

Image frames extracted from the openCV commands are processed using the Gaussian normalizer of window size $N$. Image frames are preprocessed in the 101 $\times$ 114 pixels size. Min-max classifier [8] is used to fit the pixels to be focused on the image frame. Categories of emotions to be identified are weak, good, neutral, sleepy, bored, and angry with the audio tagging [9] was also conducted by Duong et al. [6], which are carried out over here.

### *1.2.2 Ensemble-Based ELM Classification*

Given a training data set $C = \{(x_i, t_i) | x_i \in M^n,\ t_i \in M^m, \quad 1 \leq i \leq N\}$ with number of hidden nodes expressed as $H$ and activation function $f_n(x)$ and the expected learning accuracy of $\partial$. Hidden nodes are maximum allowed in the range of $H_{\max}$. $Y = f_n(M, N, H) \cdot \emptyset$ is the single hidden layer feedforward neural network. Bias [10] varies according to the number of hidden layers $H$. Hence, $H$ are the biases whereas $M$ and $\emptyset$ are weights. Ensemble algorithms often make use of bagging and boosting strategies for feature selection. Bagging determines the random sampling and feature sampling on the complete data taken as input, whereas the boosting provides the majority of the

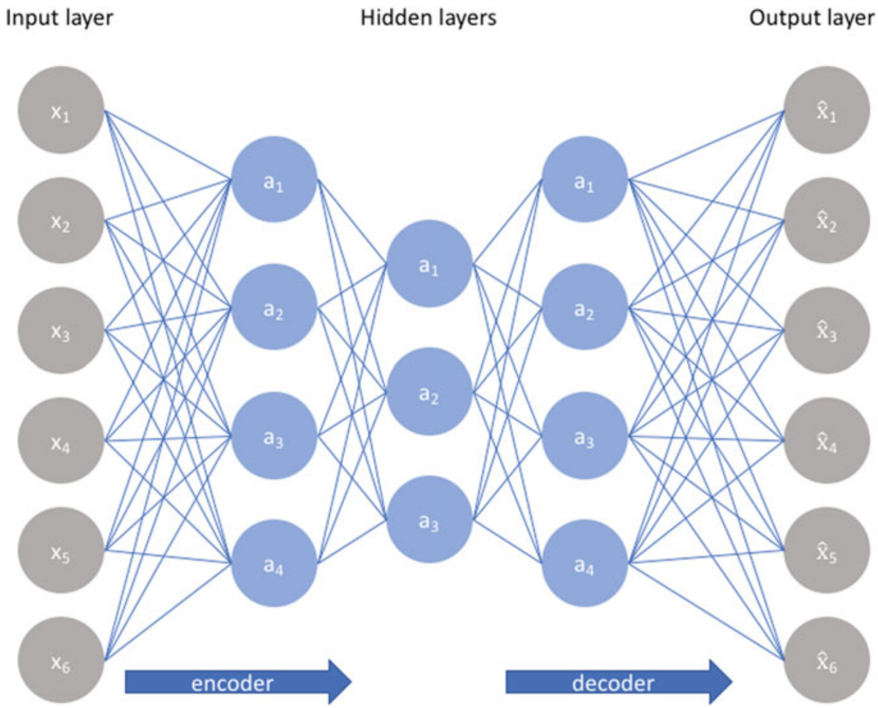


**Fig. 1.2** Simple architecture of autoencoder

votes for the classification result from the above feature selection. Accuracy is increased, whereas over fitting is reduced when ensemble algorithms are used. Iterative hidden node generation is used here for the boosting task, and extreme learning machines are constructed using the series of autoencoders. An autoencoder [11] is a type of artificial neural network. The simplest form of autoencoder is a non recurrent, feedforward neural network [12] similar to single-layer perceptron that participate in multilayer perceptron (MLP)-having an input layer, an output layer with the same number of nodes as like the input layer and one or more hidden layers. $Y = f_n(Mx_i + \emptyset)$ constitutes the output of each hidden layer. Activation function is chosen as the sigmoidal function or rectified linear unit can be chosen for $f_n$ (Fig. 1.2).

The simplest autoencoder architecture is chosen to constrain the number of nodes in hidden layers of the network. By constraining the number of nodes in the hidden layer, we can limit the information flow through the network. According to the error at the output layer identified, the model undergoes learning the bias values and weights to each node. The neural network with de-noising autoencoders is far increasingly used in the image processing area, and hence, no feature selection is needed separately to be employed on the data.

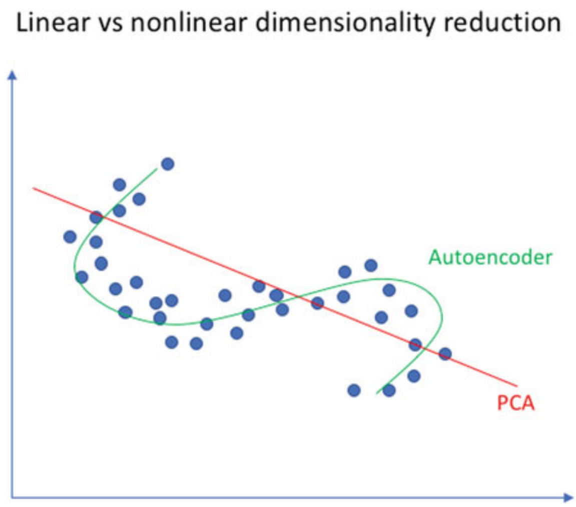


**Fig. 1.3** Linear versus nonlinear dimensionality reduction

Applications of autoencoder is listed below

1. Anomaly detection
2. Autism spectrum disorder(ASD)
3. Information retrieval
4. Drug discovery
5. Population synthesis
6. Population prediction
7. Machine translation
8. Breast cancer detection.

Autoencoder neural network follows the principle of unsupervised learning algorithm with backpropagation technique. So when backpropagation is employed, target outputs are assigned to the inputs of the neural network. The preprocessing step in autoencoders is employed with whitening process which is widely used for removing the input redundancy. It is useful in getting the pixels in the adjacent frames to be less correlated.

Autoencoders are useful in making the neural networks to learn both linear and nonlinear model relationships. PCA is employed in transforming the nonlinear model into linear model as shown below in Fig. 1.3.

## *1.2.3 Emotion Recognition System*

Given a training data set $C = \{(x_i, t_i) | x_i \in M^n,\ t_i \in M^m, \quad 1 \le i \le N\}$ with number of hidden nodes expressed as $H$ and activation function $f_n(x)$ and the expected learning accuracy of $\partial$. Hidden nodes are maximum allowed in the range of $H_{\max}$.

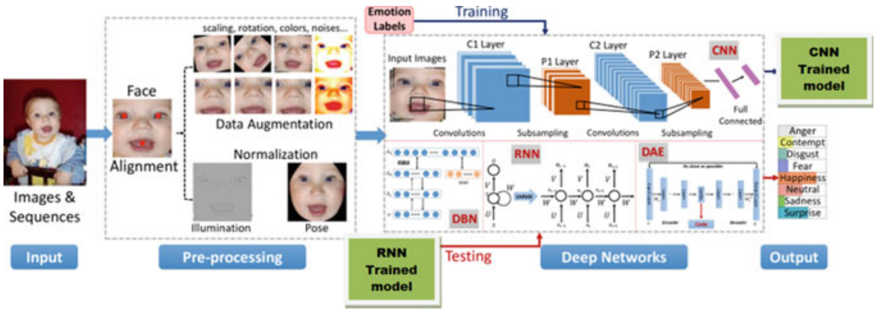


**Fig. 1.4** RNN and CNN trained emotion recognition system

$Y = f_n(M, N, H).\emptyset$ is the single hidden-layer feedforward neural network. Bias varies according to the number of hidden layers $H$. Hence, $H$ are the biases, whereas $M$ and $\emptyset$ are weights.

**Step 1**: Assign the neurons with random weights $\emptyset_1, \emptyset_2, \ldots$

**Step 2**: Assign the neurons connections with random bias $H_1$, $H_2$, …, $H_m$
**Step 3**: Input $\{(x_i, t_i) | x_i \in M^n,\ t_i \in M^m, \quad 1 \leq i \leq N\}$
**Step 4**: Calculate the hidden layer output matrix
**Step 5**: Calculate the deviation error rate obtained between the target output and the output matrix obtained in the Step 4
**Step 6**: Reduce the error $\emptyset = \underset{H}{\text{ArgMin}} \|H\beta - M\|_2^2 + \emptyset\|\beta\|_1$.

Bartlett's theory [13] states that better generalization performance can be achieved with a network with small weights. ELM classifier gives three features into existence, namely: training speed, good generalization, and universal approximation capability. EN-ELM-based ensemble classifier de-approximates the predictions of CNN. Deep recurrent autoencoders are implemented using ELM algorithm on the hidden layers of CNN (Fig. 1.4).

### *1.2.4 Preprocessing*

Gaussian normalized images are fed into the denoising autoencoder. According the limitations in autoencoder hidden layers, window size of the Gaussian local normalizer is fixed for normalization and is fed into the stacked autoencoder for larger files. Since the denoising autoencoder is having the disadvantage of sparsity, stacked autoencoder can be used alone with the Gaussian local normalizer with the window size $N$. Stacked autoencoder can be constructed by constructing a new layer before the input layer to train the input with that layer and feed it into the input

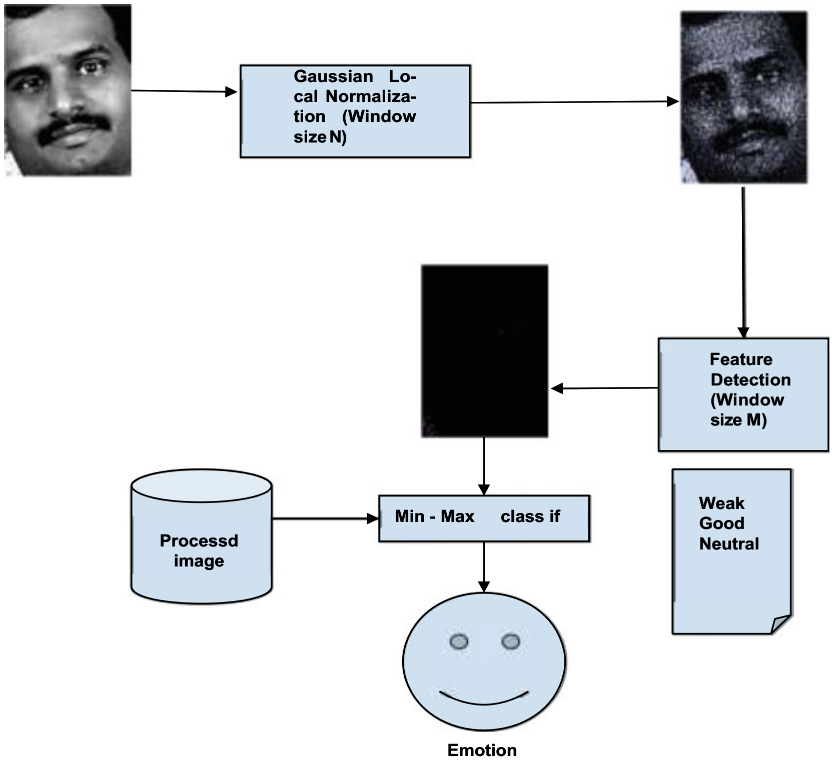


**Fig. 1.5** Outline of the emotion recognition system

layer. Stacked autoencoders contain the *n* number of sparse autoencoders in which output of each hidden layer is connected to the input of the successive hidden layers (Fig. 1.5).

Not like normal CNN, image frames alone are passed. With the help of ensemble algorithms, a temporary neuron is used to carry the time sequence from each image frame to the output of CNN, which in turn is fed to the RNN. The CNN network is used to flatten the convoluted data points into a threshold of 3.0 for moving above the fixed position. The angle of target object position is calculating with respect to the pixels of camera, and values are derived in the angle position of object. Figure 1.4 shows the generic architecture of CNN taking image frames as input and provides the sentiment text extracted according to the emotions expressed the online lecture (Fig. 1.6).

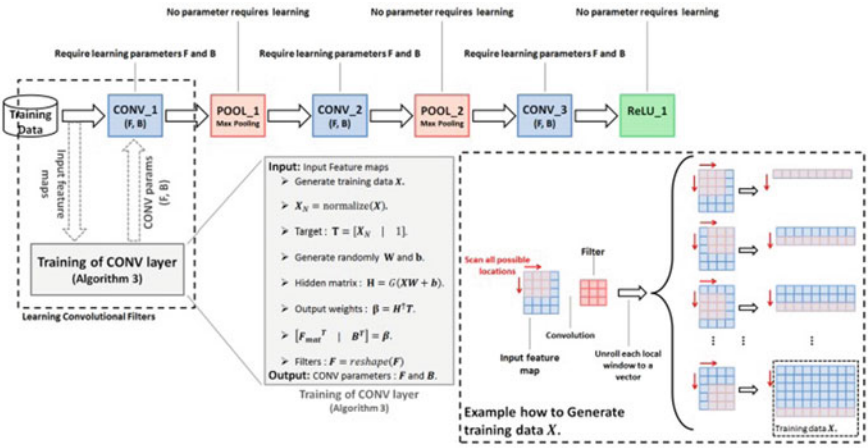


Fig. 1.6 Feature selection using CNN

## 1.3 Feature Scaling from Audio Using MFCC and HMM

Acoustic frames are constructed from the video signals using HMM. Hidden Markov models extract the NLP signals with the help Sentiment word net corpus. The Markov process simply assumes that the "future is independent of the past given present". Angry, bored, neutral, happy, sad are the hidden states to be explored. Generative sequence model according to naïve Bayes is to first decompose probability using Baye's law.

In the next step, apply the HMM word emission probability for each POS tags identified iteratively where $l$ is the length of the sentence in corpus retrieved from the video sequence by Gaussian model. Ensemble-based ELM model is used to reduce the time taken by preprocessing on video signals. Smoothing is not necessary in case of HMM transition probability from the state $l_k$ to $l_m$ is $Y_{km}$ since there are many tags. Maximum likelihood estimate is then given by

$$Y_{km} = \frac{|l_k l_m|}{|l_k|} \rightarrow$$

HMM emission probability is used for smoothening the unknown words as given below:

$$Y_{km} = \frac{|l_k l_m| + 1}{|l_k| + Q} \rightarrow$$

where $Q$ is the denotation of accepting states which are possible number of POS tags (Fig. 1.7).

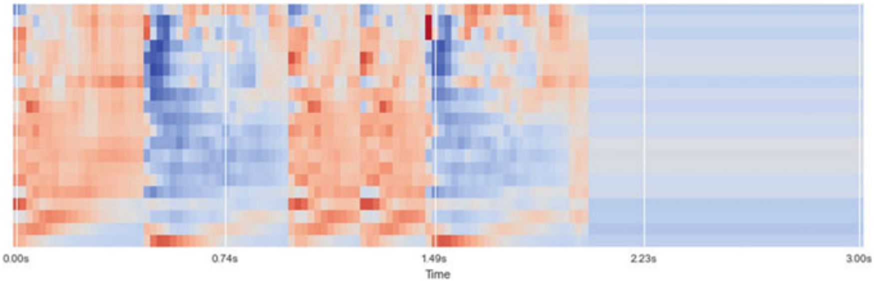


**Fig. 1.7** Feature scaling using MFCC

### *1.3.1 MFCC Frame Index*

The acoustic frames are derived from the training process of the CNN with ensemble algorithm mentioned above and LSTM is used with 10 cross fold validation. Frame-based accuracy is obtained for each acoustic information. Hori et al. in [2] used the fusion of multimodal features obtained from RNN and CNN with ensemble algorithm. 63.62% accuracy was gained using CNN ensemble algorithm and RNN with 60%. With the threshold four cross fold validation, it is observed that 72% accuracy was gained with CNN ensemble algorithm and RNN with 78.32%, whereas HMM model alone was used extracting the image transitions and sound data transitions which in turn accomplishes the 54% of accuracy alone. Gaussian models when combined with HMM model can be effectively accomplished with the accuracy of 93% (Fig. 1.8).

$$Y = f_n(Mx_i + \emptyset)$$

Sliding window framework was used for the image analysis with tweets on neural network. It is also used to ignore the noisy labels or lost labels. Positive, negative, neutral, and unsure tweets are obtained as output in sentiment analysis process in a sliding five-minute window. An online annotation tool was invented by Wang (2010) in [14]. The users of online annotation tool are allowed to submit their own opinion of the tweet. Fifty-nine percent of accuracy was achieved by [2] on the four category such as positive, negative, neutral and unsure tweets. By combining the L-Softmax and Euclidean loss in the deep learning output layer, 78.89% of accuracy was obtained, whereas L-Softmax function is used for sentiment polarity variation and Euclidean function is used for sentiment strength variation.

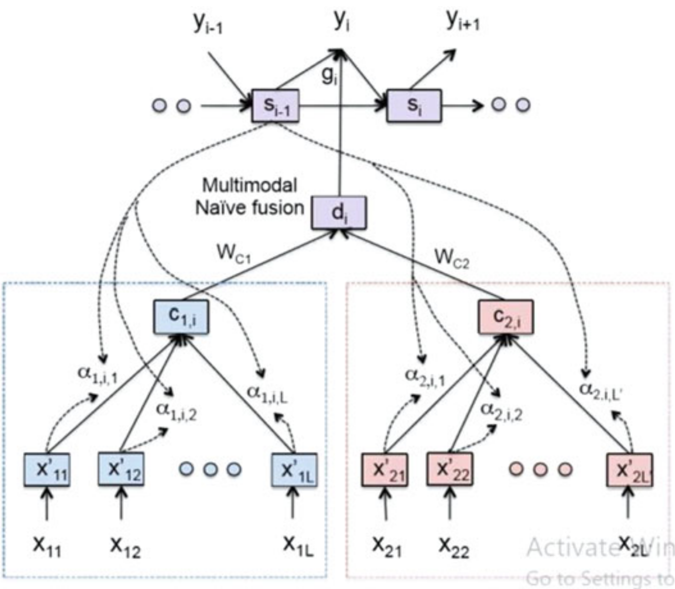


**Fig. 1.8** Multimodal features naïve fusion

## 1.3.2 *Text, Image and Audio Fusion*

Attention scheme embellished with memory networks in the deep learning architecture is adopted for memory under sentiment analysis using progressive convolutional neural network algorithm (PCNN). PCNN was proposed by Pan et al. [15] in the year 2015 high performance attention scheme for autoencoders. [16] made use of progressively trained attention shift convolutional neural networks (PTAS-CNN) and deep convolutional generative adversarial networks (DCGAN) for predicting gender from facial recognition on human facial images. AlexNet architecture is used for progressive CNN training.

By an adversarial process, two models are trained simultaneously, namely generator and discriminator neural network. Generator does the creation process and discriminator does the validation process. Generator network is updated twice for each discriminator network update for avoiding the fast convergence of discriminator. Figure 1.9 shows the fast convergence of discriminator and its validation process.

In DCGAN,

- All max pooling frames are replaced with convolutional stride.
- Transposed convolution is used for upsampling.
- Fully connected layers are eliminated.
- Batch normalization is used except the output layer for the generator and the input layer of the discriminator.