

Advanced Technologies and Societal Change

Ch Satyanarayana

Xiao-Zhi Gao

Choo-Yee Ting

Naresh Babu Muppalaneni *Editors*

Proceedings of the International Conference on Computer Vision, High Performance Computing, Smart Devices and Networks

CHSN-2020

 Springer

Advanced Technologies and Societal Change

Series Editors

Amit Kumar, Bioaxis DNA Research Centre (P) Ltd, Hyderabad, Telangana, India

Ponnuthurai Nagaratnam Suganthan, School of EEE, Nanyang Technological University, Singapore, Singapore

Jan Haase, NORDAKADEMIE Hochschule der Wirtschaft, Elmshorn, Germany

Editorial Board

Sabrina Senatore, Department of Computer and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, Italy

Xiao-Zhi Gao , School of Computing, University of Eastern Finland, Kuopio, Finland

Stefan Mozar, Glenwood, NSW, Australia

Pradeep Kumar Srivastava, Central Drug Research Institute, Lucknow, India

This series covers monographs, both authored and edited, conference proceedings and novel engineering literature related to technology enabled solutions in the area of Humanitarian and Philanthropic empowerment. The series includes sustainable humanitarian research outcomes, engineering innovations, material related to sustainable and lasting impact on health related challenges, technology enabled solutions to fight disasters, improve quality of life and underserved community solutions broadly. Impactful solutions fit to be scaled, research socially fit to be adopted and focused communities with rehabilitation related technological outcomes get a place in this series. The series also publishes proceedings from reputed engineering and technology conferences related to solar, water, electricity, green energy, social technological implications and agricultural solutions apart from humanitarian technology and human centric community based solutions.

Major areas of submission/contribution into this series include, but not limited to: Humanitarian solutions enabled by green technologies, medical technology, photonics technology, artificial intelligence and machine learning approaches, IOT based solutions, smart manufacturing solutions, smart industrial electronics, smart hospitals, robotics enabled engineering solutions, spectroscopy based solutions and sensor technology, smart villages, smart agriculture, any other technology fulfilling Humanitarian cause and low cost solutions to improve quality of life.

Ch Satyanarayana · Xiao-Zhi Gao ·
Choo-Yee Ting · Naresh Babu Muppalaneni
Editors

Proceedings of the International Conference on Computer Vision, High Performance Computing, Smart Devices and Networks

CHSN-2020

Editors

Ch Satyanarayana
Department of CSE
Jawaharlal Nehru Technological University
Kakinanda, Andhra Pradesh, India

Choo-Yee Ting
Faculty of Computing and Informatics
Multimedia University
Cyberjaya, Malaysia

Xiao-Zhi Gao 
School of Computing
University of Eastern Finland
Kuopio, Finland

Naresh Babu Muppalaneni
Department of Computer Science
and Engineering
National Institute of Technology Silchar
Silchar, Assam, India

ISSN 2191-6853

ISSN 2191-6861 (electronic)

Advanced Technologies and Societal Change

ISBN 978-981-19-4043-9

ISBN 978-981-19-4044-6 (eBook)

<https://doi.org/10.1007/978-981-19-4044-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Committee

Chief Patron

Prof. M. Ramalinga Raju, Hon'ble Vice-Chancellor, JNTUK, Kakinada

Patrons

Prof. G. V. R. Prasada Raju, Rector, JNTUK, Kakinada

Prof. B. Bala Krishna, Principal, UCEK, JNTUK, Kakinada

Honorary Chair

Prof. Allam Appa Rao, Ph.D(Engg).,D.Sc(h.c)., Former Vice-Chancellor, JNTUK, Kakinada; Chairman, IBCB, Visakhapatnam

General Chairs

Prof. Ch Satyanarayana, Registrar, JNTUK, Kakinada

Prof. S. V. Rao, Head-CSE, IIT Guwahati

Prof. Xiao-Zhi Gao, School of Computing, University of Eastern Finland, Finland

Program Chairs

Dr. Ch Sudhakar, National Institute of Technology, Warangal, Telangana

Dr. Naresh Babu Muppalaneni, National Institute of Technology, Silchar, Assam

Convenor

Dr. D. Haritha, Professor and Head in CSE, JNTUK, Kakinada

Organizing Secretary

Dr. B. N. Jagadesh, Professor and Head in CSE, Srinivasa Institute of Engineering and Technology, Amalapuram

Preface

The International Conference on Computer Vision, High Performance Computing, Smart Devices and Networks (CHSN-2020), is aimed to bring researchers together working in this area to share their knowledge and experience. In this conference, topics of contemporary interests would be discussed to provide a holistic vision on the latest technologies in computer science and engineering. The scope includes data science, machine learning, computer vision, deep learning, artificial intelligence, artificial neural networks, mobile applications development and Internet of Things, etc. Conference participants are expected to gain relevant knowledge and better understanding of the applications of computer science in various fields.

CHSN-2020 would be both stimulating and informative with the active participation of a galaxy of keynote speakers. We would like to thank all the authors who submitted the papers, because of which the conference became a story of success. We also would like to express our gratitude to the reviewers, for their contributions to enhance the quality of the papers. We are very grateful to the keynote speakers, session chairs and committee members who selflessly contributed to the success of CHSN-2020. We are very thankful to Jawaharlal Nehru Technological University, Kakinada, for providing the basic requirements to host the CHSN-2020.

Kakinanda, India
Kuopio, Finland
Cyberjaya, Malaysia
Silchar, India

Ch Satyanarayana
Xiao-Zhi Gao
Choo-Yee Ting
Naresh Babu Muppalaneni
Editors, CHSN-2020

Advisory Board

Prof. Raghu Korrapati, Waldenu University, USA
Prof. Pascal Lorenz, University of Haute Alsace, France
Prof. Viranjay M. Srivastava, University of KwaZulu-Natal, South Africa
Prof. Karthikeyan Subramanian, College of Applied Sciences, Sohar, Oman
Prof. Zhao Yang, Pacific Northwest National Laboratory, USA
Prof. Warusia Mohamed, Technical University of Malaysia, Malaysia
Prof. C. Krishna Mohan, Dean (Public and Corporate Relations), IIIT, Hyderabad
Prof. P. Srinivasa Rao, Andhra University, Visakhapatnam
Prof. K. Srinivasa Rao, Andhra University, Visakhapatnam
Prof. P. S. Avadhani, (Retd.), Andhra University, Visakhapatnam
Prof. A. Vinay Babu, (Retd.), JNTU Hyderabad
Prof. A. Govardhan, JNTU Hyderabad
Prof. L. Prathap Reddy, JNTU Hyderabad
Prof. V. Kamakshi Prasad, JNTU Hyderabad
Prof. D. Vasumathi, JNTU Hyderabad
Prof. B. Padmaja Rani, JNTU Hyderabad
Prof. A. Ananda Rao, JNTUA, Anantapuramu
Prof. V. Vijaya Kumar, Anurag Group of Institutions, Hyderabad
Prof. V. V. Krishna, Vasavi Engineering College, Hyderabad
Prof. I. Ramesh Babu, Acharya Nagarjuna University, Guntur
Mr. M. Udaya Kumar, JNTU Hyderabad
Dr. A. Chandra Sekhara Rao, IIT, Dhanbad
Dr. M. Kamala Kumari, Adikavi Nannaya University, Rajamahendravaram
Dr. R. Kiran Kumar, Krishna University, Machilipatnam

Contents

1	Character Recognition System Using CNN for Sanskrit Text	1
	R. Dinesh Kumar, M. Kalimuthu, and B. Jayaram	
2	Resampling Imbalanced Data and Impact of Attribute Selection Methods in High Dimensional Data	9
	K. Ulaga Priya and S. Pushpa	
3	Wireless sEMG Acquisition and Monitoring—A Survey	19
	R. D. Delina Rajkumari, Prakash Rajiah, A. Lakshmi Sangeetha, A. Pravin Renold, and A. Balaji Ganesh	
4	Black Hole Algorithm for BigData Anonymization	29
	U. Selvi and S. Pushpa	
5	Sleep Quality Analysis Using Motion Signals and Heart Rate	41
	R. Vijayalakshmi, Prakash Rajiah, A. Lakshmi Sangeetha, and A. Balaji Ganesh	
6	Prediction and Performance Assessment in Woman Handball Athletes by Employing Machine Learning Methods	51
	Kolla Vivek, J. Harikiran, Kunda Suresh Babu, and Ch. Anil Carie	
7	Deep Learning for Object Detection: A Survey	61
	Mohan Krishna Kotha and Kanadam Karteeka Pavan	
8	Routing Protocols: A Survey	85
	K. Raghavendra Rao and B. N. Jagadesh	
9	Performance Evaluation of Segmentation Algorithms in Non Contrast and Contrast MRI Images for Region of Interest	95
	S. Prabhu Das, B. N. Jagadesh, and B. Prabhakara Rao	
10	A Novel Curve-Energy Framework to Find the Shortest Possible Lines Through Computer Vision	113
	Chandra Sekhar Akula, Asadi Srinivasulu, and Ch. Prathima	

11	A Novel Hybrid Tracking Algorithm for Client–Server Connection Using a Machine Learning Technique	121
	P. Rama Santosh Naidu, P. Satheesh, B. Srinivas, and Venkateswarlu Sunkari	
12	Surface Damage Indicators Identification Through Simulation Studies in Computer Vision	131
	B. Srinivas, P. Satheesh, and P. Rama Santosh Naidu	
13	Gaussian Scale Concept to Reduce the Computation in Detection of Surface Defects in Machine Vision	139
	P. Satheesh, B. Srinivas, and P. Rama Santosh Naidu	
14	High-Performance Computing Framework for Virtual Memory Using CNN	147
	K. Rameshwaraiyah, S. Sree Hari Raju, and K. Ashok Kumar	
15	High-Performance Computing Center Framework for Smart Farming	157
	Chandra Sekhar Akula, Venkateswarlu Sunkari, and Ch. Prathima	
16	Generalized and Simulated Architecture for Seamless Experiment Conditions in Cloud Computing	163
	G. Hemanth Kumar, D. R. Kumar Raja, and M. Ramu	
17	Data Mining Approach for Prediction of Various Risk Factors in Supply Chain Management	173
	D. R. Kumar Raja, G. Hemanth Kumar, and P. Lakshmi Sagar	
18	Sensor-Based Radio Frequency Identification Technique for the Effective Design Process of Diverse Applications Using WSN	181
	K. Srinivasa Babu, K. Jaya Sri, and Cholleti Jyothi	
19	Reliable Smart Grid Framework Designs Through Data Processing and Analysis Process	189
	Chandra Sekhar Akula, Ch. Prathima, and Asadi Srinivasulu	
20	A Novel and Self Adapting Machine Learning Approach of ECG Signal Classification in Association with Cardiac Arrhythmia	195
	P. Rama Santosh Naidu, Golagani Lavanaya Devi, Venkata Ramana Kondapalli, and Ramesh Neelapu	

Chapter 1

Character Recognition System Using CNN for Sanskrit Text



R. Dinesh Kumar, M. Kalimuthu, and B. Jayaram

Introduction

Documents are in the form of papers that are understood by the human, but it is not possible for the computer system. So these documents are to be converted into a computer system readable form. OCR is the procedure of converting the scanned document into handwritten or printed text, symbols, letters and numerals and also can convert it into a possible format, for example, ASCII code [1]. Typically OCR can be used for processing the handwritten character and pattern recognition and is motivated greatly by a desire to enhance machine-to-man communication. Presently, few products are available for handwritten character recognition process, though different kinds of methods and procedures have been proposed [2].

Sanskrit Language

Sanskrit language is no longer spoken but still contains written text.

There are a few beliefs about the name ‘Sanskrit’:

R. Dinesh Kumar (✉)

Department of Computer Science and Engineering, Siddhartha Institute of Technology and Science, Hyderabad, India

e-mail: me.dineshkumar@gmail.com

M. Kalimuthu

Department of Computer Science and Engineering, Malla Reddy Institute of Technology, Hyderabad, India

B. Jayaram

Department of Computer Science and Engineering, RMK Engineering College, Chennai, Tamilnadu, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

1

Ch. Satyanarayana et al. (eds.), *Proceedings of the International Conference on Computer Vision, High Performance Computing, Smart Devices and Networks*, Advanced Technologies and Societal Change, https://doi.org/10.1007/978-981-19-4044-6_1

1. Sanskrit is termed as the voice of Devas.
2. Sanskrit is also termed as Devanagari because of its excessive utilization in Brahmins of Gujarat.
3. An additional viewpoint is in that Devnagar area of Kashi, therefore it was termed as Devanagari (Sanskrit).

Sanskrit is the most precise scientific basis language. For a long period, it has been Indo Aryan's script language. It is also utilized by Marathi, Hindi and Nepali languages. Widely spoken language is Hindi since its script is in Sanskrit and Sanskrit has got the dialect status. In the initial stage, Hindi was stated as the state language and Sanskrit as the start script of the few states like Haryana, Madhya Pradesh, Himachal, Uttaranchal, Bihar and so on [3]. Currently, it is found that Sanskrit is connected with every other script. In this script, all letters are equal which means there is no concept of small or capital letters and is half syllabic in nature [4, 5].

Problems in Sanskrit Text

- All separate characters are fused by a headline named 'ShiroRekha' in the case of Sanskrit script. This fusion system creates it hard to isolate separate characters from the single words.
- There are different kinds of isolated dots, which are vowel modifiers, for example, 'Chandra Bindu', 'Visarga', and 'Anuswar', which add up to the confusion.
- Descender and Ascenders recognition are also difficult, attributed to the difficult nature of language.
- It comprises composite characters.
- Minor differences in same characters.
- It comprises a huge volume of stroke and character classes.

Vowels and Consonants

Sanskrit script comprises 18 vowels ('svar') and 34 consonants ('vyanjan'). In addition, vowels and consonants are also comprised of vowel modifiers named matra's which are located at right or left part of the Sanskrit script.

Literature Survey

The framework of Sanskrit Character Recognition system is shown in Fig. 1.1.

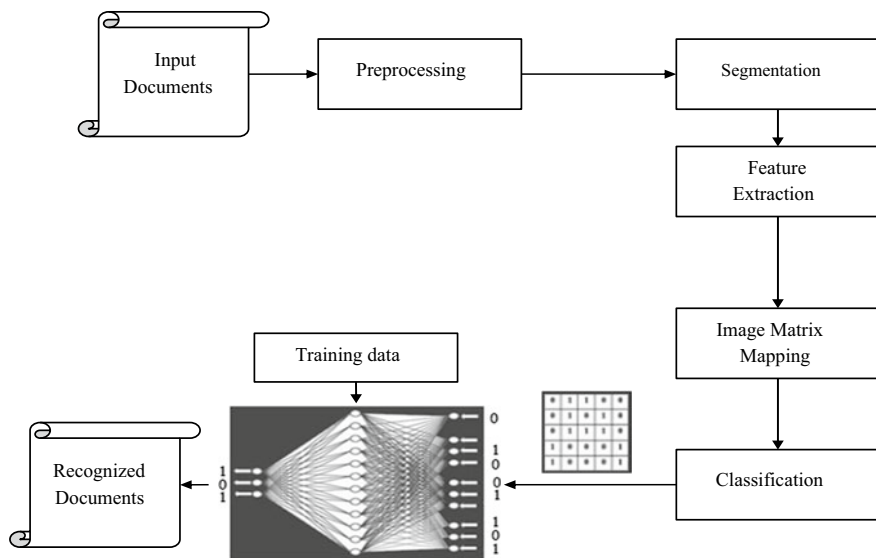


Fig. 1.1 Character recognition system

Input Documents

The documents are scanned with the proper scanner to provide the digital image for recognition. After this process, input image size is specified by the user which means the length and width of the document [6].

Preprocessing

After scanning, the next step is to remove noise from the scanned image [7]. The noise-free image is checked for skewness. Skewness is defined as the tilt (angle) in the bitmapped image of the scanned document. It is normally caused if the document is not straightly inserted into the scanner. But, most of the character recognition procedure is sensitive to the skew (orientation) of the input document, cropping is essential to implement the algorithms which can correct and detect the skew automatically.

Segmentation

After preprocessing, the noise-free image is passed to the segmentation phase, where the image is decomposed into individual characters. Sanskrit text character recognition process is carried out by applying the line and character segmentation techniques

[8]. The Sanskrit text images are segmented into the lines and each line is segmented into the words in terms of the upper modifier, consonant and lower process. Then the segmented words are converted into the straight lines which are used in Sanskrit text recognition process.

Feature Extraction

The feature extraction process examines a character segment and chooses a set of features that can be utilized uniquely to recognize the segmented text or character. The selection of a representative and stable set of a feature is the heart of the character recognition system. Different kinds of features are extracted to perform classification [9].

Classification

Classification is an important stage that is used for main decision-making process which is done by using the extracted features from the previous stages. The classification process identifies the character according to their preset rules [10]. The classification process is making decisions by using the class membership pattern. But this task is difficult because of the decision rule. Thus the feature extraction scheme is applied for reducing the misclassification probability. After this process, the classification process is done, but classification becomes an issue where characters fall into an unknown pattern.

Naïve Bayesian Classifiers

The Bayesian classifier is known to be capable of universal approximation and the output of a Bayesian network can be related to Bayesian properties. The Bayesian network has three input layers, namely input layer, hidden layer and the output layer in which each layer consumes the non-linear inputs and produces the linear output.

SVM Classifiers

The optimized features are applied to the Support Vector Machine which chooses the maximum fitness value to recognize the handwritten characters with better results. The extracted zones are considered as the features and are classified by applying the Support Vector Machine.

In this type, error function is minimized.

$$\frac{1}{2}w^r w + C \sum_{i=1}^N \xi_i \quad (1.1)$$

subject to the constraints

$$y_i(w^r \emptyset(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \dots, N \quad (1.2)$$

Note that $y \in \pm 1$ is the class label and x_i is the independent variables.

$$\frac{1}{2}w^r w - vp + \frac{1}{N} \sum_{i=1}^N \xi_i \quad (1.3)$$

subject to the constraints

$$y_i(w^r \phi(x_i) + b) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \dots, N \text{ and } \rho \geq 0 \quad (1.4)$$

CNN Classifier

CNN is a text classifier using feed-forward artificial neural networks and uses a different multilayer perceptron model to involve minimal preprocessing. The outcome of each convolution will flames when a particular pattern is identified. As a result of varying the size of the kernels and concatenating their outcome, its permits to detect patterns of various sizes containing two, three, or five adjacent words. The output patterns could be in terms like ‘I hate’, ‘very good’ and consequently CNNs can categorize them in the sentence in spite of their position.

Performance Analysis

The performance analysis of CNN with Naive Bayesian classifiers, SVM classifiers is carried out. Then the evaluated performance metrics are listed as follows (Table 1.1).

a. Sensitivity

Table 1.1 Efficiency of character recognition methods

Metrics	Naive Bayesian classifiers	SVM classifiers	CNN classifiers
Sensitivity	83.66	89.13	91.13
Specificity	84.42	90.63	92.63
Accuracy	86.56	91.05	93.45

- b. Specificity
- c. Accuracy.

Mean Square Error

MSE calculates the difference between the pixel values of the original image and the recognized image. So, the MSE is calculated by using following Equation.

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f(i, j) - f'(i, j))^2 \quad (1.5)$$

In Eq. (1.5) $f(i, j)$ is represented as the original image and $f'(i, j)$ is denoted as the recognized character image. M is the height of the image and N is the width of the image.

Unicode Mapping

The Unicode standard reflects the basic principle which emphasizes that each character code has a width of 16 bits. Unicode text is simple to parse and process Unicode characters have well-defined semantics. Hence, Unicode is chosen as the encoding scheme for the current work. After classification, the characters are recognized and a mapping table is created in which the Unicode for the corresponding characters are mapped. Based on the mapping, the Sanskrit characters are recognized with minimum error rate also enhances the recognition accuracy.

Results and Discussion

The efficiency of the CNN network is analyzed using the sensitivity and specificity metrics Table 1.2 clearly explains that the CNN method consumes minimum error rate while classifying the Sanskrit characters when compared to the existing method. The minimum error rate leads to increase the accuracy of the CNN system. From the above discussions, the proposed system classifies the Sanskrit characters with efficient manner when compared to the other traditional methods.

Table 1.2 Mean square error of CNN

Different classifiers	Mean square error
Naïve Bayesian classifier	0.9
SVM classifier	0.789
CNN classifier	0.645

Conclusion

The system explains the various processes of Sanskrit character recognition process namely preprocessing, segmentation, feature extraction and classification. The character recognition system methods are analyzed and the several quality measures such as sensitivity, specificity and accuracy have been used to analyze the effectiveness of the proposed techniques. Thus the CNN-based recognition system recognizes the exact handwritten characters in offline with minimum error rate and high accuracy when compared to the traditional classifiers. Thus, the CNN classifier overcomes the above-discussed classifiers' drawback with minimum error and high efficiency.

References

1. Agarwal, P.: Hand-written character recognition using Kohonen network. *IJCST* **2**(3), 10–18 (2011)
2. Alirezaee, S., Aghaeinia, H., Faez, K., Fard, A.S.: An efficient feature extraction method for the middle-age character recognition. *Lect. Notes Comput. Sci.* **3645**(PART II), 998–1006 (2005)
3. Ameri, M.R., Haji, M., Fischer, A., Ponson, D., Bui, T.D.: A feature extraction method for cursive character recognition using higher-order singular value decomposition. In: 2014 14th International Conference on Frontiers in Handwriting Recognition, pp. 512–516 (2014)
4. Dwivedi, N., Srivastava, K., Arya, N.: Sanskrit Word Recognition Using Prewitt's Operator and Support Vector Classification. *ICECCN*, pp. 265–269 (2013)
5. El Qacimy, B., Kerroum, M.A., Hammouch, A.: Feature extraction based on DCT for handwritten digit recognition. *Int. J. Comput. Sci. Issues (IJCSI)* **11**(6), 27–27 (2014)
6. Joshi, M.R., Sabale, M.V.V.: Offline Character Recognition for Printed Text in Devanagari Using Neural Network and Genetic Algorithm, pp. 1–8 (2013)
7. Kale, K.V., Deshmukh, P.D., Chavan, S.V., Kazi, M.M., Rode, Y.S.: Zernike moment feature extraction for handwritten Devanagari compound character recognition. *Sci. Inf. Conf.* **3**(1), 459–466 (2013)
8. Liu, X.Y., Blumenstein, M.: Experimental analysis of the modified direction feature for cursive character recognition. In: *Proceedings—International Workshop on Frontiers in Handwriting Recognition, IWFHR*, pp. 353–358 (2004)
9. Dineshkumar, R., Suganthi, J.: Sanskrit character recognition system using neural network. *Indian J. Sci. Technol.* **8**(1), 65–69. E-ISSN: 0974-5645 (2015)
10. Dineshkumar, R.: Offline Sanskrit handwritten character recognition framework based on multilayer feed forward network with intelligent character recognition. *Asian J. Inf. Technol.* **15**(11), 1678–1685. ISSN: 1682-3915 (2016)

Chapter 2

Resampling Imbalanced Data and Impact of Attribute Selection Methods in High Dimensional Data



K. Ulaga Priya and S. Pushpa

Introduction

In recent years, many real-world problems are characterized by imbalanced data where the distribution of classes is skewed. A dataset is imbalanced [1] when the classification categories are not approximately equivalent, for example one class may be extremely low (minority class) and the other class may be extremely high (majority class). This imbalanced nature of data leads the prediction algorithm to be biased toward the minority class. The poor representation of minority class affects the performance of the classification algorithm, which is evident through various assessment metrics. Applying a resampling algorithm and balancing the data resolves this issue. Studies reveal that apart from resampling technique using feature selection method improves the performance tremendously.

Current Approach

The current approach carried out in this paper for highly imbalanced data is applying feature selection technique and subsequently apply the resampling technique. The sampling technique chosen for this experiment is Random under sampling and several feature selection techniques were applied on each dataset.

K. Ulaga Priya (✉)

St. Peter's Institute of Higher Education and Research, Chennai, India

e-mail: ulagapriya@gmail.com

K. Ulaga Priya · S. Pushpa

Vels Institute of Science Technology and Advanced Studies, Chennai, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

Ch. Satyanarayana et al. (eds.), *Proceedings of the International Conference on Computer Vision, High Performance Computing, Smart Devices and Networks*, Advanced Technologies and Societal Change,
https://doi.org/10.1007/978-981-19-4044-6_2

Re-sampling Method

Several resampling technique is available in literature namely Random under Sampling (RUS), Random Over Sampling (ROS), Synthetic Minority Oversampling Technique (SMOTE), ADaptive SYNthetic sampling (ADASYN), Random Under Sampling (RUS), CONDENSED NEAREST NEIGHBOR (CNN), Edited Nearest Neighbor (ENN), and Neighborhood Cleaning Rule (NCL). This paper uses Random under sampling technique which under samples the majority data [2]. The algorithm randomly selects the majority instances which are equivalent to the number of instances in minority class. With this the class distribution can be more balanced, however, important information may be lost when examples are eliminated at random. In [3] the data is balanced and the classification would yield better results.

Feature Selection

Feature selection [4] is an important technique used in high dimensional data where irrelevant and redundant data is removed and the dimensionality is reduced. Classification will be performed with the selected attribute provided by any feature selection technique. Several feature selection techniques are available namely Information gain, Random forest selector, Lasso regression, Forward feature selection, Backward feature selection, and combination of forward and backward feature selection.

- Filter Method
 - Information gain
- Wrapper Methods
 - Forward stepwise Selection
 - Backward stepwise Selection
- Stepwise Selection
- Embedded Methods
 - LASSO
 - Random Forest

Information Gain

This is a feature selection technique using Filter method. Information gain tells us how much information [5] is given by the independent variable about the dependent variable. Information gain is helpful in case of both categorical and numerical dependent variables. For numeric dependent variables, bins are created. Information gain is the amount of information that is gained by knowing the value of the attribute