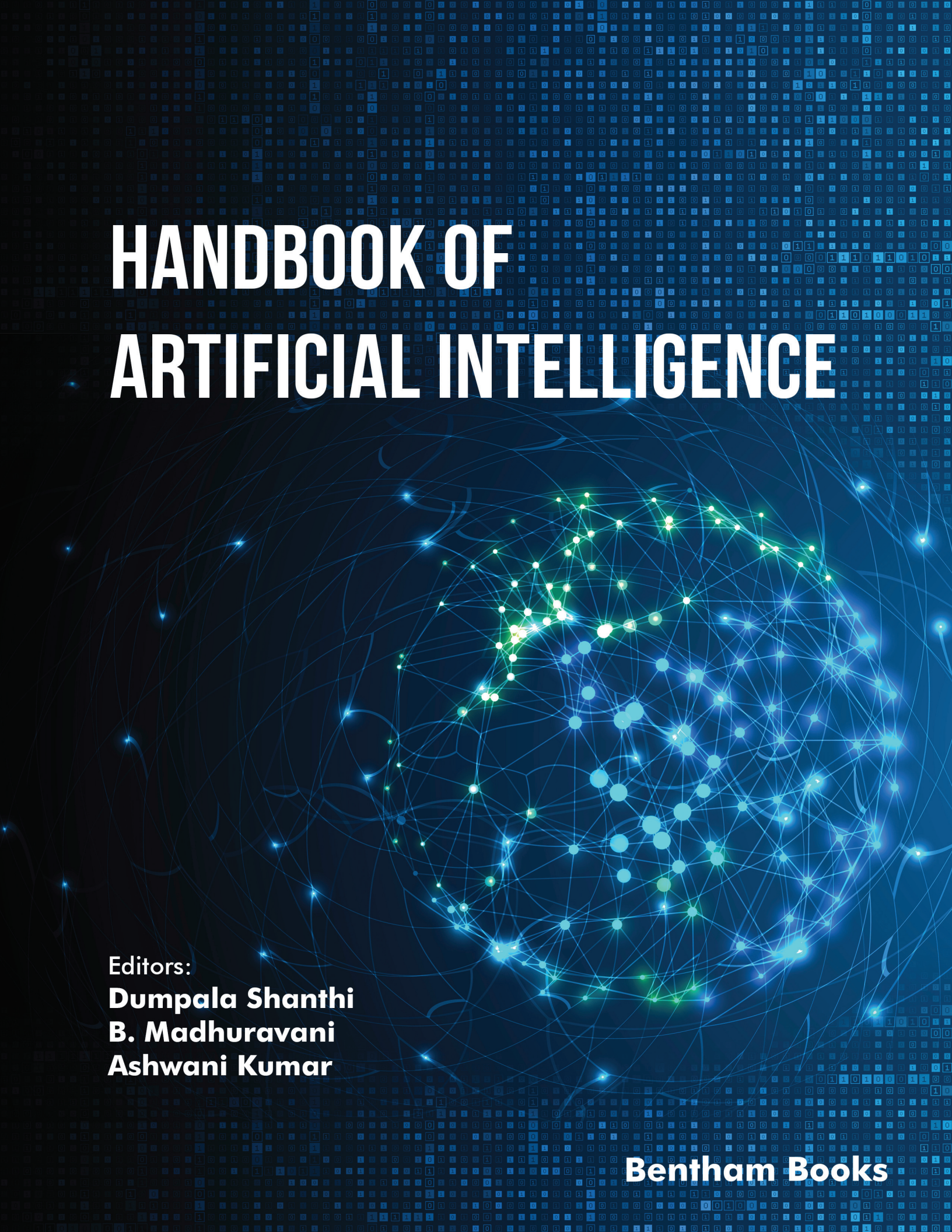
HANDBOOK OF ARTIFICIAL INTELLIGENCE



Editors:
Dumpala Shanthi
B. Madhuravani
Ashwani Kumar

Bentham Books

Handbook of Artificial Intelligence

Edited by

Dumpala Shanthi
Maturi Venkata Subba Rao Engineering College,
Nadurgul(v)
Hyderabad, Telangana
India

B. Madhuravani
Marri Laxman Reddy Institute of Engineering and
Technology, Dundigal
Hyderabad, Telangana
India

&

Ashwani Kumar
Sreyas Institute of Engineering and Technology
New Delhi
India

Handbook of Artificial Intelligence

Editors: Dumpala Shanthi, B. Madhuravani & Ashwani Kumar

ISBN (Online): 978-981-5124-51-4

ISBN (Print): 978-981-5124-52-1

ISBN (Paperback): 978-981-5124-53-8

First published in 2023.

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
80 Robinson Road #02-00
Singapore 068898
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

PREFACE i

LIST OF CONTRIBUTORS ii

CHAPTER 1 MACHINE LEARNING TECHNIQUES AND THEIR APPLICATIONS: SURVEY 1

P. Karthik, K. Chandra Sekhar and *D. Latha*

1. INTRODUCTION 1

1.1. History of AI & ML 2

1.2. Applications of ML 5

1.2.1. Speech Recognition [19] 5

1.2.2. Predictive Analytics [19] 5

1.2.3. Product Recommendation [20] 5

1.2.4. Image Recognition [19] 5

1.2.5. Video Surveillance [20] 5

1.2.6. Extraction [19] 6

1.2.7. Traffic Alerts [20] 6

1.2.8. Medical Diagnosis [19] 6

1.2.9. Sentiment Analysis [20] 6

1.2.10. Google Translate [20] 6

1.2.11. Virtual Personal Assistants [20] 6

1.3. Difference Between Traditional Programming Concepts and ML Concepts [23] 7

1.3.1. Why Must We Learn ML? 7

1.3.2. Difference Between AI & ML [23] 7

1.4. Steps to Learn ML 7

1.4.1. Data Gathering [24] 7

1.4.2. Data Preparation [24] 8

1.4.3. Selecting the Model [24] 8

1.4.4. Training the Model [24] 8

1.4.5. Evaluate the Model [24] 8

1.4.6. Parameter Tuning [24] 8

1.4.7. Make Patterns [24] 8

1.5. Types of ML 8

1.6. Basic ML Methods 9

1.7. ML in Agriculture 11

1.8. ML in Sentiment Analysis 12

1.9. ML in Stock Prediction 12

1.10. ML in Disease Prediction 12

1.11. ML in Data Mining 13

1.12. ML in COVID-19 13

1.13. ML in Cyber Security 13

1.14. ML in Cloud Computing 14

1.15. ML in Big Data Analytics (BDA) 15

1.16. ML in Recommendation System 15

1.17. Future Experiments on Real Time Problems Using ML 16

CONCLUSION 16

REFERENCES 16

CHAPTER 2 APPLICATIONS OF MACHINE LEARNING 19

K. Sudheer Babu, CH. M. Reddy, A. Swapna and *D. Abdus Subhahan*

1. INTRODUCTION 19

1.1. Machine Learning 19
1.2. Machine Learning vs. Traditional Programming 20
1.2.1. Machine Learning Work 21
2. MACHINE LEARNING ALGORITHMS 23
2.1. Introduction 23
2.2. Supervised Learning 23
2.2.1. Classification 24
2.2.2. Regression 24
2.3. Unsupervised Learning 25
3. APPLICATION OF MACHINE LEARNING 26
3.1. Augmentation 26
3.2. Automation 27
3.3. Finance Industry 27
3.4. Government Organization 27
3.5. Healthcare Industry 2 27
3.6. Marketing 28
3.7. Traffic Snarl 29
3.8. Image Recognition 31
3.9. Video Reconnaissance 32
3.10. Analysis of Sentiment 33
3.11. Recommendation of a Product 34
3.12. Online Support Using Chatbots 35
3.13. Google Translate 35
3.14. Online Video Streaming Applications 36
3.15. Virtual Professional Assistants 36
3.16. Usage of ML in Social Media 37
3.17. Stock Market Signals Utilising ML 38
3.18. Cars that are Auto-Driven 38
3.19. Dynamic Pricing- Real-Time 39
CONCLUSION 40
REFERENCES 40

CHAPTER 3 PREDICTION USING MACHINE LEARNING 45
Adluri Vijaya Lakshmi, Sowmya Gudipati Sri, Ponnuru Sowjanya and *K. Vedavathi*
1. INTRODUCTION TO MACHINE LEARNING 46
2. CLASSIFICATION OF MACHINE-LEARNING 47
2.1. Supervised Learning 47
2.2. Unsupervised Learning 48
2.3. Reinforcement Learning 49
3. BREAST CANCER PREDICTION USING ML TECHNIQUES 49
3.1. Introduction 49
3.2. Related Works 50
4. HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES 52
4.1. Introduction 52
4.2. Existing System 53
5. PREDICTING IPL RESULTS USING ML TECHNIQUES 54
5.1. Introduction 54
5.2. Related Work 56
6. PREDICTION OF SOFTWARE BUG UTILISING ML TECHNIQUE 58
6.1. Introduction 58
6.2. Related Work 59

7. PREDICTION OF RAINFALL USING MACHINE LEARNING TECHNIQUES 62
7.1. Introduction 62
7.2. Related Work 63
8. WEATHER PREDICTION USING MACHINE LEARNING TECHNIQUES 68
8.1. Machine Learning 68
8.2. Use of Algorithms 69
CONCLUSION 70
REFERENCES 71

CHAPTER 4 MACHINE LEARNING ALGORITHMS FOR HEALTH CARE DATA ANALYTICS HANDLING IMBALANCED DATASETS 75
T. Sajana and *K.V.S.N. Rama Rao*
1. INTRODUCTION 75
2. MACHINE LEARNING– AN INTELLIGENT AUTOMATED SYSTEM 76
3. TYPES OF DATASETS-BY NATURE 77
3.1. Balanced Datasets 77
3.2. Imbalanced Datasets 77
4. ISSUES WITH IMBALANCED DATASETS 78
4.1. Class Imbalance Problem 78
4.2. Classifiers Learning On Imbalanced Datasets 79
4.3. Taxonomy of Various Techniques on Imbalanced Datasets 79
5. APPLICATION OF CONVENTIONAL DATA MINING & MACHINE LEARNING TECHNIQUES FOR HANDLING CLASS IMBALANCE PROBLEM 79
6. APPLICATION OF DATA LEVEL METHODS FOR HANDLING CLASS IMBALANCE PROBLEM 81
6.1. Undersampling 81
6.2. Oversampling 83
7. APPLICATION OF ALGORITHMIC LEVEL METHODS FOR HANDLING CLASS IMBALANCE PROBLEM 85
7.1. Cost-Sensitive Classifiers 85
7.2. Ensemble Techniques 85
CONCLUSION 90
REFERENCES 91

CHAPTER 5 AI FOR CROP IMPROVEMENT 97
S.V. Vasantha
1. INTRODUCTION 97
2. GENOMICS FOR AGRICULTURE 98
3. AI FOR AGRICULTURE 99
4. AI TECHNIQUES FOR CROP IMPROVEMENT 100
5. AI-BASED CROP IMPROVEMENT MODEL (AI-CIM) 105
5.1. Automation of Modern Crop Improvement Process 105
5.2. AI Model for Enhanced Crop Breeding 105
5.2.1. Automated Selective Breeding System 105
5.2.2. Automated Plant Health Monitoring System 106
CONCLUSION 108
REFERENCES 108

CHAPTER 6 REAL-TIME OBJECT DETECTION AND LOCALIZATION FOR AUTONOMOUS DRIVING 112
Swathi Gowroju, V. Swathi, J. Narasimha Murthy and *D. Sai Kamesh*
1. INTRODUCTION 112

2. LITERATURE SURVEY 113
3. PROPOSED METHOD 117
3.1. Proposed Architecture 117
4. IMPLEMENTATION 118
4.1. Bounding Boxes 119
4.2. Anchor Boxes 120
4.3. Non-max Suppression 120
5. RELU ACTIVATION 121
6. LOSS FUNCTION 121
7. TRAINING PARAMETERS 121
8. RESULTS 122
CONCLUSION 126
ACKNOWLEDGEMENT 126
REFERENCES 126

CHAPTER 7 MACHINE LEARNING TECHNIQUES IN IMAGE SEGMENTATION 128
Narmada Kari, Sanjay Kumar Singh and *Dumpala Shanthi*
1. INTRODUCTION 128
2. LITERATURE REVIEW 133
3. METHODOLOGY 135
3.1. Collection of Data 135
3.2. Pre-processing of Images 135
3.3. Training Options 136
3.4. Define Label IDs 136
3.5. Feature Extraction 136
3.6. Feature Reduction/Selection 137
3.7. Feature Classification 137
3.8. Machine Learning 137
3.8.1. Supervised Learning 138
3.8.2. Unsupervised Learning 139
3.8.3. Reinforcement Learning 139
3.8.4. Deep Learning 140
3.8.5. Deep Reinforcement Learning 140
CONCLUSION 140
REFERENCES 141

CHAPTER 8 OPTIMAL PAGE RANKING TECHNIQUE FOR WEBPAGE PERSONALIZATION USING SEMANTIC CLASSIFIER 144
P. Pranitha, A. Manjula, G. Narsimha and *K. Vaishali*
1. INTRODUCTION 144
2. LITERATURE SURVEY 146
2.1. Challenges 147
2.2. Motivation of Research 147
2.3. Proposed Methodology 148
2.4. Generation of Web Pages 148
3. PRE-PROCESSING 148
3.1. Feature Extraction and Web Page Ranking 150
3.2. ENN-based Semantic Features 152
4. RE-RANKING OF WEB PAGES 153
4.1. Grass Hopper Optimization (GHO) 154
4.1.1. Social Interaction Calculation 154
4.1.2. Solution Updating 154

4.2. Artificial Bee Colony Algorithm (ABC) 155
4.2.1. Employed Bee Operation 155
4.2.2. Probability Calculation 155
4.2.3. Onlooker Bee Operation 155
4.3. Oppositional Grass Bee Optimization (OGBEE) 156
4.3.1. Opposition Behavior Learning (OBL) 156
4.3.2. Fitness Calculation 156
4.3.3. Updating using Grasshopper Optimization 157
4.3.4. Scout Bee Operation 158
4.3.5. Termination Criteria 158
5. RESULTS AND DISCUSSION 158
5.1. Evaluation Metrics 158
5.1.1. Precision 158
5.1.2. Recall 158
5.1.3. F-measure 159
CONCLUSION 163
REFERENCES 163

CHAPTER 9 TEXT ANALYTICS 165
Divanu Sameera, Niraj Sharma and *R.V. Ramana Chary*
1. INTRODUCTION 165
1.1. Text Analytics Basics 165
1.2. Text Analytics Examples 166
2. HOW TO GET STARTED WITH TEXT ANALYTICS 169
2.1. Analyze Your Data 169
2.2. Use BI Tools to Understand Your Data 169
2.3. Final Words 170
3. EXAMPLES AND METHODS FOR TEXT ANALYTICS 170
3.1. Text Analytics Approach 1: Word Spotting 171
3.1.1. The Simplicity of the Word-spotting Approach is What Makes it so Appealing [28] 171
3.1.2. When Word Spotting is Acceptable? 173
3.2. Text Analytics Approach 2. Manual Rules 173
3.2.1. Multiple-word Meanings Make it Hard to Create Rules 175
3.2.2. Mentioned Word! = Core Topic 175
3.2.3. Rules 175
3.2.4. Taxonomies Don't Exist for Software Products and Many Other Businesses 176
3.2.5. Not Everyone can Maintain Rules 176
3.3. Text Analytics Approach 3. Text Categorization 177
3.3.1. What is Text Categorization, and How Does it Work? 177
3.3.2. You Won't Notice Emerging Themes 179
3.3.3. Lack of Transparency 179
3.3.4. Preparing and Managing Training Data is Hard 179
3.3.5. Re-training for Each New Dataset 179
3.4. Approach 4: Topic-Modelling 180
3.4.1. What's Incredible Regarding Topic-modelling 181
3.5. Approach 5. Thematic Analysis 183
3.5.1. Thematic Analysis: How it Works 183
3.5.2. Advantages and Disadvantages of Thematic Analysis 184
3.5.3. Human in The Loop 185
4. CASE STUDY 186

CONCLUSION 189
REFERENCES 190

CHAPTER 10 HUMAN ACTIVITY RECOGNITION SYSTEM USING SMARTPHONE 195
R. Usha Rani and *M. Sunitha*
1. INTRODUCTION 195
2. LITERATURE REVIEW 196
3. MAIN TECHNIQUES 197
4. SMARTPHONE HAR COMMANDS 198
4.1. HAR Section 1: Data Cleaning Through Preprocessing 199
4.1.1. Data Filtering 200
4.1.2. Data Segmentation 200
4.1.3. Reduction 200
4.1.4. Selection of Feature 201
4.2. HAR Section II: Procedure to Perform Classification 201
CONCLUSION 202
REFERENCES 202

CHAPTER 11 SMART WATER BOTTLE WITH SMART TECHNOLOGY 204
Dumpala Shanthi
1. INTRODUCTION 204
2. EMBEDDED SYSTEM 205
3. ARDUINO NANO 206
4. PIN DIAGRAM 206
4.1. Serial Communication 206
4.2. Water Level Sensor 207
5. LDR: WORKING 208
6. APPLICATIONS 208
6.1. Light Dependent Resistor (LDR) Measurement 210
6.2. Message Management General Description 210
6.3. RA Mode 211
6.4. Tape Mode 211
6.5. Automatic Gain Control (AGC) 212
6.6. Sampling Use 212
7. MODULE 213
7.1. Types 215
7.2. Security Concerns 215
7.3. Other Devices' Interference 215
7.4. Streaming Media of Poor Quality 215
7.5. Specifications 215
CONCLUSION 218
REFERENCES 218

CHAPTER 12 REAL WORLD APPLICATIONS OF MACHINE LEARNING IN HEALTH CARE 220
Kari Narmada, Sanjay Kumar Singh and *Dumpala Shanthi*
1. INTRODUCTION 220
2. LITERATURE REVIEW 222
3. TYPES OF MACHINE LEARNING 224
3.1. Supervised Learning 224
3.2. Unsupervised Learning 224
3.3. Reinforcement Learning 224

3.4. Recommender Systems 225
4. MACHINE LEARNING APPLICATIONS IN HEALTH CARE 225
4.1. The Most Challenges For AI In Healthcare 227
4.2. Possible Risks to Generalizability in Clinical Research and Machine Learning In Health Care 228
CONCLUSION 228
REFERENCES 229

CHAPTER 13 INVESTIGATING AND IDENTIFYING FRAUDULENT BEHAVIORS OF MEDICAL CLAIMS DATA USING MACHINE LEARNING ALGORITHMS 231
Jyothi P. Naga, K.V.S.N. Rama Rao, L. Rajya and *S. Suresh*
1. INTRODUCTION 232
2. ROLE OF MACHINE LEARNING ALGORITHMS 233
3. VARIOUS KINDS OF HEALTHCARE DATA 234
4. EXISTED WORK OF DIFFERENT MODELS ON FRAUD DETECTION 235
4.1. General Model 235
4.2. Statistical Model 235
4.3. Supervised Models 236
5. MECHANISM FOR INVESTIGATING AND IDENTIFYING THE FRAUDULENT BEHAVIORS OF MEDICAL CLAIMS DATA 239
5.1. Dataset Description 239
5.2. Preprocessing 239
5.3. Model Construction 240
5.4. Mathematical Model for Experimental Analysis of Healthcare Claims 240
5.4.1. Outcomes of the Experimental Analysis 242
5.5. Methodology for Detection of Abnormal /Suspicious Claims in Healthcare Data 245
5.5.1. Outcomes of the SODAC Approach 246
5.6. Methodology for Fraudulent Behaviors of Healthcare 248
5.6.1. Outcomes Using Machine Learning Algorithms 249
CONCLUSION 250
REFERENCES 252

CHAPTER 14 SECURITY THREATS AND DETECTION MECHANISMS IN MACHINE LEARNING 255
K. Madhuri
1. INTRODUCTION 255
1.1. Supervised Learning Example 256
1.1.1. The Steps Involved in Supervised Learning 257
1.1.2. Supervised ML Algorithm Types 257
1.1.3. Regression 257
1.1.4. Classification 258
1.1.5. Supervised Learning Includes the Benefits such as 258
1.2. Unsupervised Learning 258
1.2.1. Differences between Supervised Learning and Unsupervised Learning 259
2. REGRESSION ANALYSIS IN MACHINE LEARNING 259
2.1. Outliers 260
2.2. Multi-collinearity 260
2.3. Underfitting and Overfitting 260
3. TYPES OF REGRESSION 261
3.1. Linear Regression 261
4. APPLICATIONS OF LOGISTIC-REGRESSION 262
4.1. Spam Detection 262

4.2. Credit Card Fraud 262
4.3. Tumor Prediction 263
4.4. Marketing 263
5. NEAREST NEIGHBOUR ALGORITHM (KNN) 263
5.1. Working of KNN Algorithm 263
6. MACHINE LEARNING AND SECURITY TASKS 264
6.1. Spam Detection 264
6.1.1. Content-based Filtering Technology 267
6.1.2. Case-based Spam Filtering Methods 267
6.1.3. Rules-based or Heuristic Spam Filtering Technology 267
6.1.4. Spam Filtering Technology Based on Priority Similarity 267
6.1.5. Adaptive Spam Filtering Technology 268
6.2. Limitations of Naive Bayes 269
6.2.1. Regression 270
6.2.2. Detecting Credit Card Fraud 270
6.2.3. Abuse Detection 270
6.2.4. Anomaly Detection 271
6.2.5. Advanced Fraud Detection System 272
7. SOURCE: ALTEXSOFT 272
CONCLUSION 273
REFERENCES 273
SUBJECT INDEX 275

PREFACE

Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. AI is an interdisciplinary science with multiple approaches, but advancements in machine learning and deep Learning are creating a paradigm shift in virtually every tech industry sector. Coming to machine learning, an in-built branch of AI, has changed the shape of the emerging world…!! It's not a mere exaggeration to say that ML has taken living abilities and standards a step forward. Machine learning is an application of artificial intelligence (AI) that allows systems to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of Learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow computers to learn automatically without human intervention or assistance and adjust actions accordingly.

This book covers the fundamental concepts and techniques in Artificial Intelligence, Machine Learning and Soft Computing in detail. The main focus is on real-time applications of AI, ML and Soft computing. As AICTE introduced emerging technology courses under UG Program, authors realized that there are no proper textbooks to teach students with practical orientation. The sequence of the book has been set in an appropriate manner to make it easy for students and teachers to understand concepts. The book provides insights into AI, ML and Soft computing and their applications in core areas like Agriculture, Smart Cities, Environment, *etc.*, which impact society and their vital role in the development of the nation.

Dumpala Shanthi
Maturi Venkata Subba Rao Engineering College, Nadurgul(v)
Hyderabad, Telangana
India

B. Madhuravani
Marri Laxman Reddy Institute of Engineering and Technology, Dundigal
Hyderabad, Telangana
India

&

Ashwani Kumar
Sreyas Institute of Engineering and Technology
New Delhi
India

List of Contributors

A. Manjula	Jyothishmathi Institute of Technology and Science, Telangana, India
A. Swapna	Mahatma Gandhi Institute of Technology, India
Adluri Vijaya Lakshmi	B V Raju Institute of Technology, Narsapur, Medak, India
CH. M. Reddy	Marwadi University, Rajkot, Mahender, India
D. Latha	Department of CSE, Adikavi Nannaya University, Rajamanhendravaram, India
D. Abdus Subhahan	B V Raju Institute of Technology, Narsapur, Medak, India
D. Sai Kamesh	Sreyas Institute of Engineering and Technology, India
Dumpala Shanthi	Maturi Venkata Subba Rao Engineering College, Nadurgul(v), Hyderabad, Telangana, India
Divanu Sameera	CSE Department of SSSUTMS-Sehore, MP, India
G. Narsimha	JNTUH College of Engineering, Sulthanpur, Telangana, India
J. Narasimha Murthy	Sreyas Institute of Engineering and Technology, India
Jyothi P. Naga	Department of CSE, GITAM Institute of Technology, GITAM, Visakhapatnam, A.P, India
K. Chandra Sekhar	Department of Information Technology, Maturi Venkata Subba Rao Engineering College, India
K. Sudheer Babu	Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India
K. Vedavathi	Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India
K.V.S.N. Rama Rao	Department of Computer Science & Engineering, KLH, Hyderabad, India
K. Vaishali	Jyothishmathi Institute of Technological Sciences, Karimnagar, Telangana, India
K. Madhuri	Anurag University, Hyderabad, India
M. Sunitha	CVRCE, CSE, Hyderabad, India
Niraj Sharma	Department of CSE, SSSUTMS-Sehore, India
Narmada Kari	Amity University Rajasthan, Jaipur Rajasthan, India
P. Karthik	Department of Information Technology, Maturi Venkata Subba Rao Engineering College, India
Ponnuru Sowjanya	School of Technology GITAM(Deemed to be University), Rudraram, Hyderabad, India
P. Pranitha	Jyothishmathi Institute of Technology and Science, Telangana, India
R.V. Ramana Chary	Department of IT, B V Raju Institute of Technology, Narsapur, Medak Telangana., India
R. Usha Rani	CVRCE, CSE, Hyderabad, India
Rajya Lakshmi	Department of Computer Science and Engineering, JNTUK UCEV, Vizianagaram, A.P., India

Sowmya Gudipati Sri	School of Technology GITAM (Deemed to be University), Rudraram, Hyderabad, India
S.V. Vasantha	CSE Department, Vardhaman College of Engineering, Hyderabad, India
Swathi Gowroju	Sreyas Institute of Engineering and Technology, India
Sanjay Kumar Singh	Amity University Rajasthan, Jaipur Rajasthan, India
S. Suresh	Neerukonda Institute of Technology & Sciences, Sangivalasa, Bheemunipatnam, Vishakapatnam, Andhra Pradesh, India
T. Sajana	Department of Computer Science & Engineering, K LEF, Guntur, India
V. Swathi	Sreyas Institute of Engineering and Technology, India

CHAPTER 1

Machine Learning Techniques and their Applications: Survey

P. Karthik[1,*], K. Chandra Sekhar[1] and D. Latha[2]

[1] *Department of Information Technology, Maturi Venkata Subba Rao Engineering College, India*

[2] *Department of CSE, Adikavi Nannaya University, Rajamanhendravaram, India*

Abstract: Machine Learning has become part of our daily life directly or indirectly. Several Machine Learning techniques are being used in several areas to increase the effective usage of computers by human beings. Over the past few decades, the concepts of AI and ML have rapidly increased their importance in various application areas, and a lot of research has taken place and is still in progress. Even then, there is a lot more to explore regarding the applications of AI and ML in our day-to-day lives. The reason to do so is only to improve the existing process of work in various areas and also to give valuable suggestions and scope for further research. In this context, here we tried to discuss basic concepts related to machine learning and gave a brief overview of various application areas of machine learning.

Keywords: Machine Learning (ML), Artificial Intelligence (AI), Cloud Computing, Big Data Analytics.

1. INTRODUCTION

Artificial Intelligence (AI), the most prominent research area of computer science is basically concerned with the development of machines that can think and work similarly to human beings. AI is an interdisciplinary research area with many sub-branches like Artificial Neural Networks (ANN), Machine Learning (ML), *etc*. ML, which is said to be a sub-division of AI, is based on the concept of developing machines that can learn from data and formulate patterns from it with very less interference from human beings. Basically ML allows machines to learn from data and improve themselves to solve tasks.

[*] **Corresponding author P. Karthik:** Department of Information Technology, Maturi Venkata Subba Rao Engineering College, India; E-mail: karthik_it@mvsrec.edu.in

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

1.1. History of AI & ML

In the year 1950, Alan Turing developed the Turing test. The objective of conducting the Turing test was to check whether a machine can think like a human or not. Even though this was a basic level test, but it later had a huge impact on AI [1, 2].

The Turing test [3, 4] can be defined as a simple question-and-answer game played between a human and a machine. The person who asks questions is also a machine. Now the task of the machine is to check and analyze the answers given, and analyze whether the machine is human or not. Even though Alan Turing foresaw that by the 21st century, machines will be able to think and act like humans, it didn't happen completely like now. For example, if we consider any kind of Interactive Voice Response System (IVRS) used in mobiles or ATMs *etc*., we can easily identify it as a machine but not a human. That is how we could clearly differentiate between a machine and a human.

In the year 1952 [5, 6], Arthur Samuel developed the first computer program that was able to learn. The program was the game of checkers, and the IBM computer improved at the game, the more it played, studying which moves made up winning strategies and involving that list of moves inside its program.

In 1957 [7], the first neural network for computers was designed by Frank Rosenblatt to simulate the thought process of the human brain.

In the year 1960 [8, 13], MIT developed a Natural Language Processing program to act as a therapist. The program was called ELIZA, and it was a successful experiment. But it was still using scripting to do its work. It was a key milestone for the development of a subset of Machine Learning called Natural Language Processing (NLP) which is being widely used even today.

In the year 1967 [9, 11, 12], the "nearest neighbor" algorithm was written, which allows computers to start using very basic pattern recognition. Later it was used to form a route map for traveling salesmen, who starts at a random city but ensure they visit all cities in the tour K-Nearest Neighbor or KNN is one of the important and famous Machine Learning algorithms.

In the year 1970 [10, 13], the concept of Backpropagation was developed. Back Propagation is nothing but a set of algorithms used in Deep Learning. They modify deep learning neural networks in such a way that they can correct themselves.

In the year 1979 [12], the students of Stanford University developed a "Stanford cart" that has the ability to navigate the obstacles within a particular room on its own.

In the year 1980 [13], Artificial Neural Network (ANN), which is a multilayered neural network, was developed by Kunihiko Fukushima.

In the year 1981 [12], the concept of Explanation Based Learning (EBL) was introduced by Gerald Dejong. EBL enables a computer system to analyze the given data and formulate a rule that can be followed by discarding data that is not so important. This was a very early stage of developing the concept of a Machine Learning Algorithm.

In the year 1985 [12], Nettalk was invented by Terry Sejnowski and Charles Rosenberg. NETtalk is an automated procedure that was developed to learn how to pronounce the English text written, when text is shown as input and matching phonetic transcriptions for comparison. The purpose of creating NETtalk was to identify the methods of correctly learning to pronounce the text in English.

In the year 1989 [13], the Reinforcement Learning concept was developed. In reinforcement learning, artificial intelligence faces a situation similar to that of a game where it learns to achieve a goal in a potentially complex and uncertain environment. To make the machine to act according to the needs of the programmer, the artificial intelligence gets either a benefit or a fine for the actions it performs. It aims to maximize the total benefit. Here, the programmer will specify game rules but won't give any clue to the machine regarding the rules of the game or the procedure for solving the game. It is up to the machine to decide and formulate how to win the game by following the trial and error method, thereby increasing its own rewards.

In the early 1990's [12], the work on Machine Learning shifted to a "data-driven approach" from a "knowledge-driven approach". From here, experts have started developing computer programs, and those computers must analyze or gain knowledge from that data.

In 1995 [13, 14], Random Forest Algorithm and Support Vector Machines algorithm were developed. These are considered to be some of the most important Machine Learning algorithms.

In 1997/98 [13, 15], Sepp Hochreiter and Jürgen Schmidhuber introduced LSTM. Long Short-Term Memory (LSTM) [16] networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a required behavior in domains like machine translation, speech

CHAPTER 2

Applications of Machine Learning

K. Sudheer Babu[1,*], CH. M. Reddy[2], A. Swapna[3] and **D. Abdus Subhahan[4]**

[1] *Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India*

[2] *Marwadi University, Rajkot, Mahender, India*

[3] *Mahatma Gandhi Institute of Technology, India*

[4] *B V Raju Institute of Technology, Narsapur, Medak, India*

Abstract: In this chapter, we briefly discuss various real-time applications of machine learning algorithms. Machine Learning Algorithms explain the following topics: Introduction to ML algorithms, Supervised Learning, Classification, Regression (**Linear Regression, Logistic Regression, Decision Tree, Naive Bayes, Support Vector Machine, Random Forest, AdaBoost, Gradient-Boosting Trees),** and Unsupervised Learning (**K-Means Clustering, Gaussian Mixture Model, Hierarchical Clustering, Recommender Systems, PCA/T-SNE).** Application of Machine Learning explains various real-time applications like augmentation, automation, finance, government, healthcare, marketing, traffic alerts, image recognition, video surveillance, sentiment analysis, product recommendation, online support using chatbots, Google translate, online video streaming applications, virtual professional assistants, machine learning usage in social media, stock market signals using machine learning, auto-driven cars, and real-time dynamic pricing.

Keywords: Machine-Learning, Supervised learning, Un-Supervised learning, Naive bayes, Support vector machine, Random Forest, AdaBoost, Gradient-boosting.

1. INTRODUCTION

1.1. Machine Learning

Machine-learning is a set of computer algorithms that are able to learn from experience and improve themselves without any necessary direct programming.

One part of AI is machine learning, which employs mathematical methods to combine data and predict a result that may be used to make decisions [1, 2].

[*] **Corresponding author K. Sudheer Babu:** Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India; E-mail: sudheerkuntumalla@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

This new idea is based on the concept that a computer could learn from a dataset (for example) in order to come up with proper results on its own. Machine learning is closely related to both data mining and Bayesian predictive modelling. The computer takes information as input and uses an algorithm to come up with answers [3].

Providing a recommendation is a common machine-learning problem-for example, entire proposal of films or series depending on the client's history of information for those with a Netflix account. Unsupervised learning is being used by tech businesses to improve the customer experience by personalising recommendations [4].

1.2. Machine Learning *vs.* Traditional Programming

Traditional programming (Fig. **1**) is very different from machine learning. In conventional programming, a programmer writes all of the instructions with the help of an expert in the field for which the software is being made. Every instruction is based on something logical, and the computer will follow the logical assertion, and make something happen. As the system gets more intricate, more rules need to be written. It can quickly become impossible to keep up [5, 6].

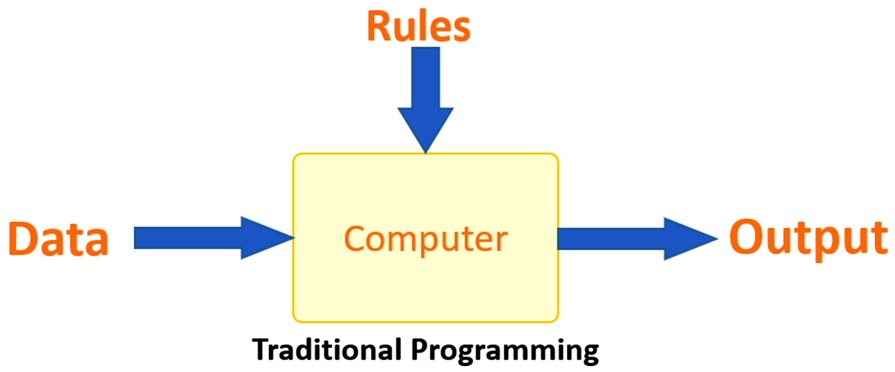


Fig. (1). Traditional Programming.

This problem is meant to be solved *via* machine learning. The machine deduces the relationship between the input/output information and creates the instructions. Each time there is fresh data, the programmers do not need to design new rules. To improve efficacy over time, the algorithm adjusts in reaction to novel information and involvements [7, 8].

1.2.1. Machine Learning Work

We will now study how Machine Learning (ML) works:

The "brain" of the machine learning system is where all learning takes place. Learning occurs in exactly the same way on the computer as it does in people. Experience is the single most important teacher for humans. When we have more knowledge at hand, it is much simpler to make intelligent predictions. By analogy, the chances of succeeding in an unknown circumstance are lower than the chances of succeeding in a known situation. A similar instructional method is used for machines. In order to generate an accurate forecast, the machine investigates a specific example. When we provide the machine with a scenario that is comparable to what might occur, it is able to make accurate predictions [9]. On the other hand, just like a person, if the computer is given a scenario that it has never encountered before, it will struggle to accurately forecast what will occur in the future.

As per the (Fig. **2**) learning and inference are the two main goals of machine learning. First and foremost, the machine acquires by perceiving designs. The information helped in making this finding. One of the most important tasks for a data researcher is to painstakingly select the information to provide to the computer [10, 11]. A feature vector is a collection of attributes used to solve an issue. A feature vector can be considered a subset of information that will be utilised in solving the issues.

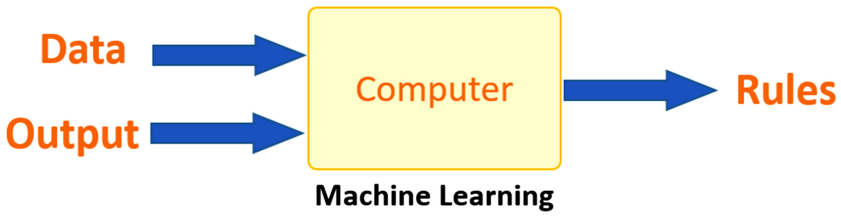


Fig. (2). Machine Learning.

The computer employs a few sophisticated algorithms to improve realism and turn this finding into a model. As a result, the step in learning is utilized to define and summarise the information in order to create a model.

Consider an example; shown in Fig. (**3**) the device which attempts to comprehend the connection between a person's pay and his or her possibility of dining at a posh restaurant. It turns out that the system detects a link between pay and dining at a very good quality establishment: It is referred to as model Inferring [12, 13].

CHAPTER 3

Prediction using Machine Learning

Adluri Vijaya Lakshmi[1,*], Sowmya Gudipati Sri[2], Ponnuru Sowjanya[2] and **K. Vedavathi[3]**

[1] *B V Raju Institute of Technology, Narsapur, Medak, Telangana, India*

[2] *School of Technology GITAM(Deemed to be University), Rudraram, Hyderabad, India*

[3] *Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India*

Abstract: This chapter begins with a concise introduction to machine learning and the classification of machine learning systems (supervised learning, unsupervised learning, and reinforcement learning). 'Breast Cancer Prediction Using ML Techniques' is the topic of Chapter 2. This chapter describes various breast cancer prediction algorithms, including convolutional neural networks (CNN), support vector machines, Nave Bayesian classification, and weighted Nave Bayesian classification. Prediction of Heart Disease Using Machine Learning Techniques is the topic of Chapter 3. This chapter describes the numerous heart disease prediction algorithms, including Support Vector Machines (SVM), Logistic Regression, KNN, Random Forest Classifier, and Deep Neural Networks. Prediction of IPL Data Using Machine Learning Techniques is the topic of Chapter 4. The following algorithms are covered in this chapter: decision trees, naive bayes, K-Nearest Neighbour Random Forest, data mining techniques, fuzzy clustering logic, support vector machines, reinforcement learning algorithms, data analytics approaches and Bayesian prediction techniques. Chapter Five: Software Error Prediction by means of machine learning- The AR model and the Known Power Model (POWM), as well as artificial neural networks (ANNs), particle swarm optimisation (PSO), decision trees (DT), Nave Bayes (NB), and linear classifiers, are among the approaches (K-nearest neighbours, Nave Bayes, C-4.5, and decision trees) Prediction of Rainfall Using Machine Learning Techniques, Chapter 6: The following are discussed: LASSO (Least Absolute Shrinkage and Selection Operator) Regression, ANN (Artificial Neural Network), Support Vector Machine, Multi-Layer Perception, Decision Tree, Adaptive Neuro-Fuzzy Inference System, Wavelet Neural Network, Ensemble Prediction Systems, ARIMA model, PCA and KMeans algorithms, Recurrent Neural Network (RNN), statistical KNN classifier, and neural SOM Weather Prediction Using Machine Learning Techniques that includes Bayesian Networks, Linear Regression, Logistic Regression, KNN Decision Tree, Random Forest, K-Means, and Apriori's Algorithm, as well as Linear Regression, Polynomial Regression, Random Forest Regression, Artificial Neural Networks, and Recurrent Neural Networks.

[*] **Corresponding author Adluri Vijaya Lakshmi:** B V Raju Institute of Technology, Narsapur, Medak, Telangana. India; E-mail: vijaya.a@bvrit.ac.in

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

Keywords: Reinforcement learning, Convolutional Neural Networks, Weighted Nave Bayes Classifier, Fuzzy clustering logic, Linear Classifiers k-Nearest Neighbours.

1. INTRODUCTION TO MACHINE LEARNING

ML is a rapidly evolving technology that automatically allows computers to learn from previous data. Machine learning employs a variety of algorithms to create mathematical models and make predictions based on past data or knowledge. It can be used to identify pictures, recognize speech, filter emails, auto-tag in Facebook, recommend systems and many more activities [1, 2].

This machine learning course introduces ML and the many ML techniques available, including supervised, unsupervised and reinforcement learning. Regression and classification models, clustering approaches, hidden Markov models, and other sequential models will all be covered [3].

According to (Fig. **1**), a Machine Learning system learns from previous data, constructs prediction models, and predicts the result whenever fresh data is received. The amount of data helps to construct a better model that predicts the output more precisely, hence the accuracy of anticipated output depends on the amount of data [4].

If we have a complex situation for which we need to make predictions, rather than writing code for it, we may just input the data to generic algorithms, and the machine will develop the logic based on the data and forecast the outcome. Machine learning has shifted our perspective on the issue [5, 6]. The following block diagram depicts how the ML algorithms work.

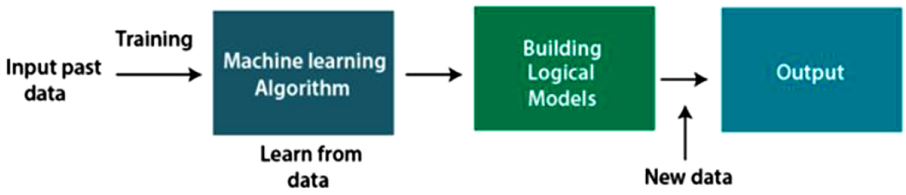


Fig. (1). ML prediction model.

2. CLASSIFICATION OF MACHINE-LEARNING

ML can be categorized into three types on a general level mentioned in Fig. (**2**):

1. Supervised learning

2. Unsupervised learning

3. Reinforcement learning

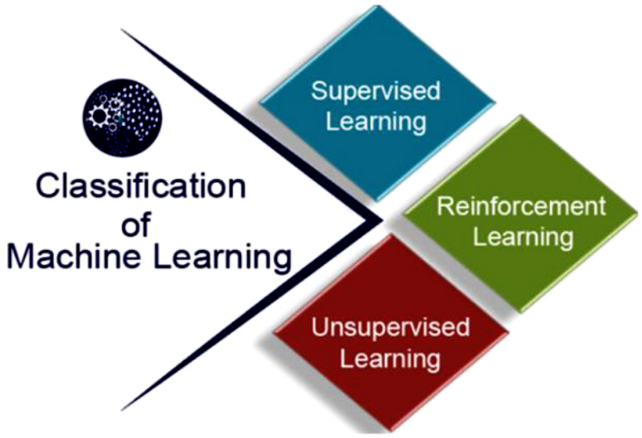


Fig. (2). Machine-Learning Classification.

2.1. Supervised Learning

It can be defined as a form of ML method in which the ML system sample labelled data is provided to train it on, and it then predicts the result. To interpret a set of data and to acquire details about each data, the system constructs a model using labelled data in Fig. (**3**). After the training and processing are complete, the model can be tested by supplying a sample dataset to see if it accurately predicts the output [7, 8].

The goal of supervised learning is to plot the input into output data. It depends upon the supervision, as well as it is similar to learning by a student under the guidance of a teacher. Spam filtering is one example of supervised learning.

CHAPTER 4

Machine Learning Algorithms for Health Care Data Analytics Handling Imbalanced Datasets

T. Sajana[1,*] and **K.V.S.N. Rama Rao**[2]

[1] *Department of Computer Science & Engineering, K LEF, Guntur, India*

[2] *Department of Computer Science & Engineering, KLH, Hyderabad, India*

Abstract: In Machine Learning, classification is considered a supervised learning technique to predict class samples based on labeled data. Classification techniques have been applied to various domains such as intrusion detection, credit card fraud detection, *etc.* However, classification techniques on all these domains have been applied to balanced datasets. Balanced datasets are those which contain equal proportion of majority and minority examples. However, in real-time, obtaining balanced datasets is difficult because majority of the datasets tend to be imbalanced. Developing a model for classifying imbalanced datasets is a challenge, particularly in the medical domain. Accurate identification of a disease-affected patient within time is critical as any misclassification leads to severe consequences. However the imbalanced nature of most of the real-time datasets presents a challenge for most of the conventional machine learning algorithms. For the past few years, researchers have developed models using Conventional machine learning algorithms (linear and nonlinear) are stating unsatisfactory performance in classifying imbalanced datasets. To address this problem of skewed datasets several statistics techniques & robust machine Learning techniques have been developed by the researchers. The discussion on handling imbalanced datasets in the healthcare domain using machine learning techniques is a primary focus of this chapter.

Keywords: Healthcare, Skewed distribution, Machine learning techniques, Imbalanced datasets.

1. INTRODUCTION

Now-a-days, people are affected by various kinds of diseases and diagnosing them has become a real challenge. Especially in the case of vector-borne diseases like malaria and dengue, the disease should be diagnosed immediately; otherwise, it may even lead to the death of the patient [1].

[*] **Corresponding author T. Sajana:** Department of Computer Science & Engineering, K LEF, Guntur, India; E-mail: sajana.cse@kluniversity.in

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

Hence, in diagnosing these types of diseases, some level of expertise is required. Pathologists are specialized professionals who analyze medical data samples and diagnose the disease. Pathologists need to assess data thoroughly before concluding the patient's status. In several cases, it was observed that there are variations among pathologists' conclusions, *i.e.*, one pathologist concludes that the patient has a disease, and other pathologists may predict that the patient is not affected by the disease [2]. There was a conflict among decisions on the same data. These variations are due to the following reasons:

- Human vision/ bias causing errors
- Lack of expertise
- Manual analysis complications.

As a result, an automated system is necessary to minimise the drawbacks brought about by such manual diagnosis errors [2, 3]. The Director-General of the World Health Organisation (WHO), Dr. Tedros Adhanom Ghebreyesus, indicated that the usage of computer-based models in the medical arena is producing accurate findings in analysing the patient situation [4]. The training of automated systems is quite good, which results in better outcomes. These better results are helpful for complex data processing, which in turn assists clinicians [5].

2. MACHINE LEARNING– AN INTELLIGENT AUTOMATED SYSTEM

To analyze complex data with computer-aided models, an expert artificial system is required. Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed [6]. Machine learning focuses on the development of computer programs that can access and learn from data. But the primary goal is to allow computers to learn automatically without human assistance. The learning process basically starts with data collection or observations [5]. Examples include learning to diagnose patients in the medical field, learning to drive an autonomous car, learning to identify new astronomical structures, and so on. Methods or algorithms for learning by machines are often categorized into supervised learning, unsupervised learning, semi-supervised learning and Reinforcement learning. The classification of disease-affected patients is the primary focus of this work, which uses supervised learning. Supervised learning is also known as classification, in which predictions can be done based on the labeled examples in the dataset [7].

3. TYPES OF DATASETS-BY NATURE

By nature, datasets can be typically categorized into two types: balanced and imbalanced. Balanced Datasets are those whose class samples ratio is equally proportionate to each other, whereas, in imbalanced datasets, one class sample ratio dominates the other class sample ratio *i.e*., the class samples ratio are not proportionate to each other [8].

3.1. Balanced Datasets

Balanced datasets are those whose class samples ratio is equally proportionate to each other. A well-balanced dataset is suitable for the classifiers to perform classification efficiently.

Example: Consider the aircraft dataset as shown in the Table **1**.

Table. 1. Aircraft classification dataset.

Wing type	Engines	Nose	Intake	Fuselage	Class
Mid	Low	Flat	Nose	Cigar	Foreign
Mid	Low	Flat	Nose	Sleek	Foreign
Low	Low	Snub	Nose	Sleek	Foreign
High	Medium	Point	Body	Thick	Domestic
High	low	point	Body	Thick	Domestic
Low	Medium	Point	Body	Sleek	Domestic

From Table **1**, the aircraft dataset has of 6 tuples, which are defined with attributes-wing type, engines, nose, intake and fuselage. The class label for the dataset is defined with two values – foreign and domestic. From the dataset, 3 tuples belong to one class called “foreign" and the remaining 3 tuples are belongs to the “domestic” class. The ratio of Examples of both classes is equally proportionate to each other. Hence, the dataset is called a balanced dataset.

3.2. Imbalanced Datasets

Datasets in which one class sample ratio dominates the other class sample ratio is called imbalanced dzataset [9].

Example: Let us consider the Website Phishing dataset, which consists of 10 parameters as shown below in Table **2**.

CHAPTER 5

AI for Crop Improvement

S.V. Vasantha[1,*]

[1] *CSE Department, Vardhaman College of Engineering, Hyderabad, India*

Abstract: The introduction of high-performance genomic technologies into plant science has resulted in the generation of huge volumes of genomic information. Moreover, for biologists to deal with such complex, voluminous dataand infer some significant findings in order to improve crop quality and quantity has presented a big challenge to them. The advent of Artificial Intelligence (AI), Machine learning (ML) and Deep Learning (DL), facilitated automated tools for more efficient and better analysis of the data. Another crucial process that needs to be automated in field farming is the timely and precise diagnosis of crop diseases which plays a vital role in the prevention of productivity loss and reduced quantity of agricultural products. ML provides a solution to solve these problems by automatic field crop inspection. Recently, DL techniques have been widely applied for processing images to obtain enhanced accuracy. This chapter describes the need of AI in Agri-Genomics; it also includes various contemporary AI solutions for the Crop Improvement process and presents the proposed AI-based Crop Improvement Model (AI-CIM).

Keywords: Plants, Genomics, Crop improvement, Artificial intelligence, Machine learning, Deep learning.

1. INTRODUCTION

Engineering the crop for the benefit of humankind, is as aged as the agricultural science itself. Around decamillennium ago, our ancestors made a shift from foraging or hunting to the cultivation of crops for their survival. Due to this shift, the continuous program on the improvement of crops that serves as a staple food has begun.

To meet up the demands of an ever-increasing population, dynamically changing environmental factors, and scarce resources for agriculture, crop improvement has become inevitable for humankind [1]. Therefore, the goal of plant science is to boost the productivity and quality of the crop and, at the same time, to maintain

[*] **Corresponding author S.V. Vasantha:** CSE Department, Vardhaman College of Engineering, Hyderabad, India; E-mail: s.v.vasantha@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

the ecological system for sustainable agriculture [2]. Majorly, two methods are being practiced for improving the crops, according to Lawrence, a plant biologist at Harvard University. The first method is Selection, which is based on the plant's genetic sequence variation. Early farmers selected plants with beneficial traits, such as one that bore large-sized fruits or one that was easy to harvest. This selection of plants or seeds is used for subsequent year's crop cultivation. The traditional method of natural selection is being strengthened by artificial selection.

The second method is breeding, which is employed to develop new varieties of plants, usually called cultivars, that have desired traits by crossing the selected parent plants. It involves the production of progeny that possibly have traits of both the parent plants after selecting each parent with a preferred trait. But this process was a miss or hit for early farmers due to the lack of awareness of genetic propagation of traits and even being unable to predict the outcome traits of a specific cross until 1900. Mendel, a renowned biologist, has thrown light on plant inheritance and deduced that every single trait, such as colour, is taken from the two parent genes, one from each parent [3].

Plant Breeders select plants with desired traits or characteristics, for example, plants with greater yield and resistance to major diseases or better-adapting capability to varied environments, based on performance statistics, plant pedigree and genetic information. Traditional breeding techniques may take around ten years to produce a new variety of crops from the same kind of species [1]. With the advent of Genomics and Genetic Engineering along with AI techniques in agriculture, crops are drastically improved, and several different cultivars are being produced in various species.

2. GENOMICS FOR AGRICULTURE

Application of Genomics knowledge to the field of agriculture is coined as Agri-Genomics as depicted in functions and respondig. (Fig. **1**), which is one of new emerging fields in plant science with an objective to increase the productivity as well as sustainability of the crops and livestock.

Involvement of Plant Genomics in agriculture has a significant role in the improvement of crops over the conventional breeding process [4]. Genomics technologies have contributed significantly to modern agriculture by forming the basis for a better interpretation of how a plant functions and respond to the prevailing environment and attaining the desired objectives in breeding programs to improve the quality, disease resistance and productivity of the crops. Genetic markers related to key traits have enabled knowledge-based breeding, which has reduced the traditionally time-consuming process of an extensive selection of plants from the field [5].

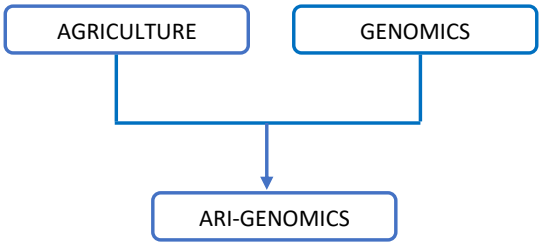


Fig. (1). Agriculture and Genomics into Agri-Genomics.

Usage of marker-assisted recurrent selection technique for maize crop, a 3% yield raise per cycle under the water stress circumstances was reported [6]. For legumes, authors [7] reported a 26.5kg /ha/year genetic gain by examining the historic data over 80 years of soybean farming.

Genome editing technologies, such as CRISPR/Cas, permit the intended mutation of nearly any plant genome sequence producing a novel variant with expected traits of interest [8]. Genomic selection (GS) presents the opportunity to boost staple crop production in reduced time with intended traits. The main task of GS in Genomics deals with the prediction of Breeding-Values Estimated from the Genome, *i.e.*, GEBVs for each sample available with its genotype (Genes) data [9, 10].

3. AI FOR AGRICULTURE

Modern-day agriculture employs the usage of Argi-Genomics along with Artificial Intelligence (AI) techniques to automate the processing of huge and complex genomic data generated for the betterment of the crop yield, which is depicted in Fig. (**2**).

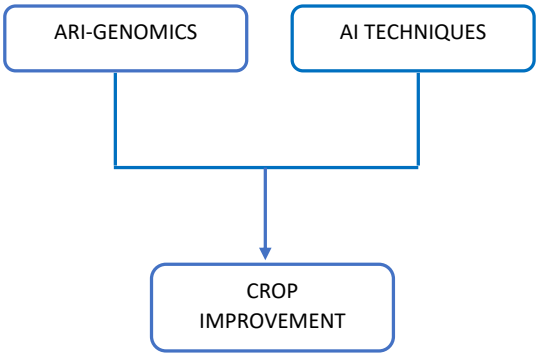


Fig. (2). Modern Crop Improvement Process.

CHAPTER 6

Real-Time Object Detection and Localization for Autonomous Driving

Swathi Gowroju[1,*], V. Swathi[1], J. Narasimha Murthy[1] and **D. Sai Kamesh[1]**

[1] *Sreyas Institute of Engineering and Technology, Hyderabad, Telangana 500068, India*

Abstract: The term "object detection" refers to a technology that enables humans to recognise specific types of things present in visual media. One of the important applications of the technique is autonomous driving cars. In the application, the activity is to detect the various objects present in the single image frame. Examples of objects belonging to multiple classes are trucks, bikes, persons, cars, dogs, and cats. For this task, we use object localization and classification as we have to locate multiple objects in the image. Various techniques available in the market based on Deep Learning use inbuilt architectures such as VGG-16 and InceptionV3. Using these techniques to solve the problem is a reasonable solution but the response time from these architectures may not be feasible as the autonomous vehicles have to react in less than 0.02 milliseconds in order to avoid collisions of all sorts. So using YOLO, we simply predict the classes and the bounded co-ordinates of the object in a single run of the model and detect multiple objects from the image rather than focusing only on the interested regions of the image as formerly employed by various models. YOLO is fast and accurate with the help of Convolution Neural Networks and is less likely to produce localization errors.

Keywords: Object localization, Autonomous driving, Convolution neural networks.

1. INTRODUCTION

We know that autonomous driving or self-driving cars is a topic that has been in the news lately. The integral part of the self-driving application is object detection using a camera that captures nearly 40 to 80 frames per second. Once the image is captured, meaning a frame is under process, the image should be fed to the mainframe for detection and classification, which means object localization should be done. We call the process PERCEPTION, which is simultaneously detection, prediction, localization, mapping *etc*.

[*] **Corresponding author Swathi Gowroju:** Sreyas Institute of Engineering and Technology, Hyderabad, Telangana 500068, India; E-mail: swathigowroju@sreyas.ac.in

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

There may be many challenges because the control system should receive the detection results quickly, say in milliseconds, and the process should be repeated multiple times in 5 seconds. Hence, the car adjusts its acceleration and lane system. If too many parameters are present, the system becomes complex, and the time taken to classify increases. To undo this mistake, CNN comes into the picture.

2. LITERATURE SURVEY

Gene Lewis explains the use of SimpleNet architecture [1], which is a near-rea--time model. The architecture is capable of producing results at a processing rate of 11 frames per second with an mAP accuracy of 12.83% on the validation set. OverFit [2, 3] is a CNN model which performs object detection, recognition and localization. It is the most successful model, which is eight layers deep and depends on the overlapping bounding box scheme, where it iteratively merges them to the high-probability predictions. VGG16 [3] is an attempt to work better than OverFit with a 16-layer and 19-layer architecture with functionality similar to OverFit Table **1**.

Table 1. Literature Survey.

Author	Methodology Used	Accuracy (%)	Dataset Used	Classes Used	Time taken for Prediction
S. Gidaris *et al*. [4]	MR_CNN_MORE_DATA	93.4	Own dataset with 5000 images	20	0.05 sec
R. Girshick *et al*. [5]	R-CNN VGG BB	91.4	Own dataset of images	20	0.053 sec
J. Dong *et al*. [6]	BabyLearning	91.7	Own datasets	25	-
S. Gidaris *et al*. [7]	MR CNN S CNN	90.4	Own dataset	30	-
S. Ren [8]	NoC	83.3	Own dataset	50	0.043 sec
R. B. Girshick *et al*. [9]	R-CNN BB	73.3	Own dataset	40	-
B. Hariharan *et al*. [10]	SDS	70.7	Own dataset	20	-

Vehicle detection and its statistical information are considered significant for tracking vehicles on highway and identifying them through an aerial view. With many of the cameras on the highways and on the roads, monitoring and keeping surveillance is highly needed in obtaining information through traffic video

footage, *etc*. When viewing the vehicle at a certain angle, it is mandatory to install the camera to cover greater distances. But it is a risky dataset for the researchers as the object becomes smaller when the distance between the camera and the vehicle increases. If the illumination is not proper on the images, the detection rate will fall down [11].

Currently vision-based vehicle detection is divided into two types, namely, traditional-based machine vision methods and deep learning-based machine vision methods. Traditional methods generally use the motion of the vehicle to separate it from the static background and fixed background. Such as background subtraction, change in the frame difference, and optical flow method. This method is best suited for moving vehicles. The optical field represents the direction of the pixel and also the momentum of the pixel.

In the video frame difference method, the variance of two or three consecutive frames is calculated from the pixel intensities to check the background. The foreground is separated by using the thresholding method. This method works well for stopped or parked vehicles. Each frame image is compared with the background and suppressing noise, we can also detect the moving vehicle. When each frame is compared to the background model, we can segment the moving object.

Scale In-variant Feature Transform(SIFT) and SURF methods are also used for detecting moving vehicles on the road. Using the 3D ridges, the correlation curves were detected to classify the vehicles.

In spite of the scarcity of availability of areal images through satellite, it is always a challenging task to find the vehicles from the images. Due to the low resolution and complexity of surrounding images, there are numerous algorithms that were proposed for detection, specifically on a real image. These models can be categorized into two types. One is object modelling based algorithms. Another is Object appearance-based algorithms. The algorithms such as R-CNN, YOLO models basically depend upon the quality of the input image. On the other hand, object appearance-based models such as MAD and Histogram Oriented Gradient (HOG) feature extraction models work with statistical information of the predicted object. HOG method, clubbed with SVM(Support Vector Machine), is used to identify the location of the vehicle.

CNN (Convolutional Neural Networks) are state-of-art methods nowadays due to the use of deep convolution neural networks; the CNNs have achieved success in every field of computing and vision processing. The ability to learn in CNN has shown an amazing success rate in object detection, too. They can perform different tasks, such as segmentation and classification, using various algorithms.

CHAPTER 7

Machine Learning Techniques in Image Segmentation

Narmada Kari[1,*], **Sanjay Kumar Singh**[1] and **Dumpala Shanthi**[2]

[1] *Amity University Rajasthan, Jaipur, Rajasthan, India*

[2] *Maturi Venkata Subba Rao Engineering College, Hyderabad, India*

Abstract: Image is an important medium to express information easily. This paper deals with the content of image segmentation with machine learning. Segmentation is the process of extracting the information required from the image. Machine learning is the process that helps to classify to obtain good results. A number of algorithms are designed for the segmentation process. The algorithms are selected based on the application. Quality segmentation can be applied if the algorithm is fixed at the application level. Standalone methods can be used for real-time applications. Schematic segmentation is one of the best techniques used for segmenting images. Machine learning combines basic techniques to produce good results. The algorithms vary for different input images like MRI, CT Scans, Colour images, *etc*. Algorithms like k-mean clustering are mostly used in processing. Many problems occur in segmentation which can be removed by Bayesian architectures. The usage of machine learning improves accuracy and efficiency. Labeling, training and testing are some of the methods used in segmentation through machine learning.

Keywords: Image segmentation, Feature classification, Machine Learning types.

1. INTRODUCTION

Information is an important means of communication; the information is of various types, of which images are one of the mediums.

Image is a two-dimensional medium. Over the entire image, the information exists only at some part of the image [1].

Therefore, instead of processing the entire image, the informative part of the image can be separated and processed. This separation process is called image segmentation. This is an important process in the image processing field [2].

[*] **Corresponding author Narmada Kari:** Amity University Rajasthan, Jaipur, Rajasthan, India;
E-mail: narmadakari@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

Image segmentation is the primary step in all applications. Various segmentation techniques are developed, but there will always be a gap in the research for the segmentation process. Up to now, there is no suitable technique for all the input images. Relative measures are considered for computing the values. There is a difference between computer-generated results and human-generated results. The human eye cannot see the sharpness in the segmented process. The system also easily finds minute errors, so the simulation is important to segment the image [3]. Machine learning helps to increase the accuracy rate to the point where the exact information extracted is visualized. Images are basically of three types:

- Black and white images (Binary images)
- Greyscale images
- Color Images

The researcher chooses any one type for their segmentation. Generally, medical images are greyscale images. Operating greyscale and color images are difficult compared to binary images. The segmentation method is compared to wide applications instead of narrow applications. Good performance segmentation is not an easy task to generate [4].

Exploiting the process of image segmentation is important. In most of the applications, it is easily viewed whether the segmentation is flexible or inferior to the process. Segmentation is calculated by maximum human-aided techniques, which the researchers mostly use. Human-aided is perfect in performance but takes more time to compute and is tedious. The combination of segmentation with machine learning is always a great process for researchers. Standalone segmentation is also equally good for accuracy. Segmentation is estimated with the characteristics of the image. Reference images are not needed for any type of segmentation.

In real-time, the image is captured, and automatic segmentation can be done on the image with the help of machine learning. If the system is very big, then the segmentation is used in pre-processing of an image. Pre-processing deals with

- Amplification
- De-noising
- Filtering
- Smoothening
- Sharpening
- Segmentation

The pre-processing improves the quality of an image and increases to obtain a good quality of segmentation. The basic segmentation of an image is shown in Fig. (**1**).

Fig. (1). Image segmentation.

Generally, many techniques are performed by the simulated algorithm for segmentation in the system. Around 20 different techniques are performed with segmentation, and all the results are stored in the system [5]. Many number of segmentation techniques are available; some of the important techniques are given below.

- Thresholding
- Histogram-based bundling
- K means clustering
- Region growing
- Watersheds
- Active contours
- Conditional Markov random fields
- Graph cuts
- Sparsity
- Schematic
- Instance

The main objective of this paper is to determine the required area through the segmentation process with the help of advanced segmentation algorithms. The

CHAPTER 8

Optimal Page Ranking Technique for Webpage Personalization Using Semantic Classifier

P. Pranitha[1,*], A. Manjula[1], G. Narsimha[2] and **K. Vaishali[3]**

[1] *Jyothishmathi Institute of Technology and Science, Telangana, India*

[2] *JNTUH College of Engineering, Sulthanpur, Telangana, India*

[3] *Jyothishmathi Institute of Technological Sciences, Karimnagar, Telangana, India*

Abstract: Personalized webpage ranking is one of the key components in search engines. Moreover, most of the existing search engines focus only on answering user queries, although personalization will be more and more important as the amount of information available on the Web increases. Even though various re-ranking algorithms are developed, providing prompt responses to the user query results in a major challenge in web page personalization. Therefore, an efficient and effective ranking algorithm named the Oppositional Grass Bee optimization algorithm is developed to re-rank the web documents in the webpage personalization system. The proposed algorithm is designed by integrating the Oppositional Grass Hopper (OGHO) and Artificial Bee Colony optimization (ABC) algorithms. The concept of fictional computing and the foraging behavior realize the re-ranking process more effectively in the web environment. However, the semantic features extracted from the web pages make the process more effective and achieve optimal global solutions through the fitness measure. The proposed OGBEE Ranking algorithm effectively captures and analyzes the ranking scores of different search engines in order to generate the re-ranked score result.

Keywords: Page ranking, Artificial bee colony optimization, Web mining, Grass hopper optimization, Feature Extraction, Web Personalization, Factor based Features, Rank-based features, Hyperlink-based ranking, Hyperlink induced topic search, ENN-based Opposition Behavior Learning.

1. INTRODUCTION

A search engine is a piece of software that enables users to enter a query and transmit it along with other relevant websites relevant to the information they are looking for. There are countless online papers on the WWW (*i.e.*, on every page

* **Corresponding author P. Pranitha:** Jyothishmathi Institute of Technology and Science, Telangana, India; E-mail: ppranitha7@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

of the website), some of which may be found on various parts of the site with a multitude of information, many sophisticated steps that need to be done by the search engine in order to acquire information from the Internet [1].

Several search engines follow the following sequence of operations: searching, indexing, sorting, and data ranking. Although search engines are helpful for obtaining useful material on the internet, they are important tools for discovering the latest information. All queries, processing and ranking, are performed through the use of search and indexing techniques without accessing the source code of web pages. An important approach to boost search rankings is by automating web pages based on the subject of a user's interest [2]. Bespoke experiences are created by leveraging acquired information, such as the user's navigational activities and navigational patterns, to alter the overall design of a site. It is a method for websites that includes data on how people use the site as well as structural data to rank and provide suggested web page links to visitors. Web Personalization permits the modification or priority delivery of content depending on individual users' explicit or implied preferences. Details supplied either by the user or the implicit/derived behavior of the user and preferences, such as links clicked, or pages seen, are the priority or page ranking for a certain content or web page [3].

One of the most successful strategies when it comes to looking for query phrases is rating the query word. Web pages may be categorized into those that are navigation-oriented and those that include top-notch content [4]. The ranking algorithms in use now were designed within the context of the words and phrases that existed at the time. Page-Rank is the most well-known approach for ranking websites. Although previously mentioned, the amount of connections connecting websites is key to the Page-Rank algorithm [5]. The Page-Rank algorithm assigns thousands of web-based websites to a page ranking score on the World Wide Web [8]. The Hyperlink Threaded Search (HITS) algorithm and Page-Rank algorithm are excellent web-ranking and crawling techniques [6]. The number of hyperlinks pointing to an inbound link or web page is taken into consideration by the hyperlink-based ranking algorithm, which is used early in the search process [7].

In contrast to the link-based ranking strategy, which is used to discover websites related to the specific page query, the content-based ranking approach aims to return the web pages that are relevant to the query. The queries the user enters are paired with keywords found in online logs and web documents that the user maintains, thereby enabling these algorithms to work effectively (fairy tale) [8, 9].

This project aimed to build a website customization re-ranking technique that used the OGBEE algorithm. Pre-processing involves the web page's design, scripting, and optimizing the webpage for search engines, followed by feature

extraction and re-ranking. Next, the question is sent to Google and Bing search engines, which are responsible for returning relevant web pages for the user. Before processing, all the pages are delivered to the number of pages received. Then, noise is eliminated from the pages before they are supplied to the machine [10]. Also, the HTML file is processed to become a text file beforehand. In the features extraction stage, essential properties of online documents are extracted and made available. Finally, the approach that was proposed is utilized to re-rank the competitors. The major contribution of this research is explained as follows:

- **Proposed Ranking Algorithm for Personalization:** For re-ranked websites in search engines, an efficient and optimum web page ranking web is created. The ranking of the web pages allows the re-ranking of measurements. The re-rank measure analyses the distance from the query to the online sites and is the lowest possible answer [11].

2. LITERATURE SURVEY

A number of different existing page ranking approaches are tested to discover whether they provide benefits and pitfalls, with the hope of coming up with a new methodology. Vijaya P and Chander S [1] developed a new method to re-rank the websites, and this algorithm produced a new rating for the sites called the "lion rank." This technique located and obtained the documents that were needed. While the results were favorable, the performance of this approach was subpar. Makkar A and Kumar N [2] created a user behaviour rating system using the learning automata approach. This method reduced computing costs and energy consumption while improving page rank scores for websites. This method did not work for detecting spam pages. Desarkar M.S *et al*. [3] created the unsupervised rank algorithm in order to minimize the number of poor rankers. This product not only surpassed the current market standard, but also was capable of meeting the unique meta-search criteria. It was created by Philip L.H. and Xu H [4] to maximize the achievable expectations of websites. As the rankings increased exponentially, and cognitive classification levels were established, the classification system analyzed the degree of cognitive categorization. Different cognitive rankings and origins did not make them a good group. The multi-layer perceptron (MLPNN) neural network created by Banaei H and Honarvar A.R [5] incorporates machine learning to evaluate this website. With increased precision, performance improved. The neural network's settings were more challenging to set. To determine page rank, You K *et al*. [6] created a Markovian Kaczmarz method. When dealing with spamming nodes, this method worked well. In a short time, it produced many linkages, but the dynamics of opinion were not integrated into social networks. A dual model based on the algorithm of the rank page was

CHAPTER 9

Text Analytics

Divanu Sameera[1,*], Niraj Sharma[2] and **R.V. Ramana Chary[3]**

[1] *CSE Department of SSSUTMS-Sehore, MP, India*

[2] *Department of CSE, SSSUTMS-Sehore, India*

[3] *Department of IT, B V Raju Institute of Technology, Narsapur, Medak, Telangana, India*

Abstract: This chapter covers text analytics definitions, how to get started with text analytics, examples and approaches, and a case study. The chapter gives examples of existing text analytics applications to show the wide range of real-world implications. Finally, as a guide to text analytics and the book, we give a process road map. Chapter 2 (How to Get Started with Text Analytics) briefly explains the Analyse Your Data, Use BI Tools to Understand Your Data and Final Words. Chapter 3 (Examples and Methods for Text Analytics) explains various Text Analytics Approaches 1: Word Spotting Text Analytics Approach 2. Manual Rules Text Analytics Approach 3. Text Categorization Approach 4: Topic Modelling Approach and 5. Thematic Analysis. Applications of Word Spotting Text Analytics Approach, Manual Rules, Text Categorization Approach, Topic Modelling Approach and Thematic Analysis are discussed with real-time examples. Chapter 4 discusses the case study, the following real-time application, Word Cloud Explorer, to illustrate its analytic capabilities.

Keywords: Text Analytics, Word Spotting Text Analytics, Thematic Analysis, Text Categorization.

1. INTRODUCTION

1.1. Text Analytics Basics

Text analytics is the act of automatically converting an enormous amount of indefinite text into numerical data in order to identify insights, trends and patterns. This technique, when combined with data visualisation tools, allows businesses to comprehend the story behind the numbers and make better decisions [1 - 3]. You may have come across comparable ideas like text mining and text analysis while

[*] **Corresponding author Divanu Sameera:** Department of SSSUTMS-Sehore, MP, India;
E-mail: ameerach24@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

looking for a definition of text analytics. Therefore, before we go into the details, let's revise the key distinctions between these phrases [4].

Text mining, text analysis, and text analytics are the terms that refer to the process of examining disorganized text so as to extract insights. Text analytics, however, gathers these results and turns them into something that can be quantified and represented through charts and reports, whereas text mining (or text analysis) provides qualitative insights [5].

Text analysis and text analytics are frequently used in tandem to provide a comprehensive knowledge of all types of text, including emails, social media posts, surveys, and customer service tickets, among other things. For example, text analysis technologies have major use in determining how people feel about a company on social media (sentiment analysis) or in comprehending the major issues in product evaluations (topic detection) [6]. Text analytics, else ways, uses the results of text analysis to discover patterns, such as a spike in negative comments, and gives you actionable insights that you can use to enhance your product, such as addressing a defect that is causing your users to be frustrated [7].

1.2. Text Analytics Examples

Among the most important intriguing applications of text analytics, customer feedback analysis is primarily in business. This could entail assessing product and service reviews to determine how your consumers rate your business, processing the findings of open-ended customer survey replies, or monitoring what customers are saying about your brand on social media [8, 9].

Let's imagine you want to look at open-ended NPS replies to see which components of your business are being discussed by your customers.

The first step would be to perform a topic analysis, which is a text analysis approach that tags NPS replies based on predetermined categories like Feature Requests and Customer Service. Listed below are a few examples:

Feature Request: “I wish you had the option to export the data.”

“Customer service is extremely helpful and friendly!” Customer Service

After you've categorised each NPS response, you can use text analytics to find patterns and insights throughout the entire dataset, and then display the results in charts or reports. Recently employed text analytics to map common subjects in NPS replies, as shown in the graph below:

According to (Fig. (**1**)) , the majority of negative comments are about product features, while the majority of positive comments are about product usability and customer support. According to the negative feedback, adding popular feature requests could help improve customer happiness [10].

Fig. (1). Response Tags Analysis.

Customer service is another fascinating application of text analytics in business, in addition to customer feedback analysis [11]. Text analytics, for example, found major use in examining the content of support requests to better understand your customers' needs, motivations, and expectations, as well as providing suggestions on how to rethink your customer experience approach [12, 13].

Text analytics can also be utilised to extract crucial data from various customer support channels like as email, chat, and social media, in addition, to support tickets shown in Figs. (**2** & **3**). For example, we looked at telecommunication business customer service conversations on Twitter and discovered that T-Mobile has the most favourable interactions [14, 15].

CHAPTER 10

Human Activity Recognition System Using Smartphone

R. Usha Rani[1] and **M. Sunitha**[1,*]

[1] *CVRCE, CSE, Hyderabad, India*

Abstract: Recognition of human activity has a wide range of applications in medical research and human survey systems. We present a powerful activity recognition system based on a Smartphone in this paper. The system collects time series signals with a 3-dimensional Smartphone accelerometer as the only sensor, from which 31 features in the time domain and frequency domain are created. The quadratic classifier, k-nearest neighbor algorithm, support vector machine, and artificial neural networks are used to classify activities. Feature extraction and subset selection are used to reduce dimensionality. In addition to passive learning, we use active learning techniques to lower the cost of data tagging. The findings of the experiment demonstrate that the categorization rate of passive learning is 84.4 percent and that it is resistant to common cell phone postures and poses.

Keywords: Activity Recognition, Gyroscope, Segmentation.

1. INTRODUCTION

The demand for understanding human activities has increased in the health-care domain, particularly in elder care support, rehabilitation assistance, diabetes, and cognitive disorders [1]. A significant amount of resources can be saved if sensors can assist caregivers in constantly recording and monitoring patients and reporting automatically when any abnormal behavior is detected. The study has benefited other applications, such as a human survey system and a location indicator. Many studies have successfully identified activities using wearable sensors with very low error rates, but the majority of previous work has been done in laboratories with very constrained settings [2, 3].

Although readings from many body-attached sensors produce a low error rate, the sophisticated configuration is not practical in reality. In our research, we employed low-cost, commercially accessible cell phones as sensors to detect

* **Corresponding author M. Sunitha:** CVRCE, CSE, Hyderabad, India; E-mail: m.sunitha@cvr.ac.in

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

human activity. Smartphone's' increasing popularity and processing capacity make them a perfect choice for non-intrusive body-attached sensors. According to US mobile subscriber statistics, around 44 percent of mobile users in 2011 had Smartphones, and 96 percent of these smartphones had built-in inertial sensors such as an accelerometer or gyroscope [4, 5]. According to research, gyroscopes can aid in activity identification, however, their contribution alone is not as effective as accelerometers [6, 7]. Asgyroscopes are not as widely accessible in cell phones as accelerometers; our method exclusively employs 3-dimensional accelerometer measurements. Unlike many others, we removed the limitations of connecting a sensor to a set body position with a fixed device orientation, unlike much previous research. In our concept, the phone may be positioned anywhere around the waist, including the jacket, pocket, and jeans pocket, with arbitrary orientation. These are the most prevalent postures in which individuals use their cell phones [8].

When a new activity is added to the system, a training procedure is always necessary. Due to sensor variation, parameters of the same algorithm may need to be learned and changed when the algorithm runs on various devices. However, labeling time-series data takes time, and it is not always viable to ask users to label all of the training data. As a result, we advocate for the use of an active learning strategy to expedite the training process. Given a classifier, active learning intelligently searches the unlabeled samples and learns the parameters based on the right labels returned by the oracle, which is generally a human. In this manner, users label just the examples requested by the algorithm, reducing the total number of training samples necessary. To the best of our knowledge, no previous research has been conducted on applying active learning to the problem of human activity identification. In this paper, we created a lightweight and accurate system for recognizing human behaviors on a Smartphone. Furthermore, active learning algorithms are being developed to minimize labeling time and stress. We determine the optimal learning algorithm for our system in terms of efficiency and accuracy by testing and comparing several learning algorithms.

2. LITERATURE REVIEW

For years, experts have researched human activity recognition and presented several solutions to the problem. Existing techniques usually employ a vision sensor, an inertial sensor, or a combination of the two. The use of machine learning and threshold-based algorithms is common. Machine-learning algorithms are often more accurate and dependable, but threshold-based methods are quicker and simpler. Body posture has been captured and identified using one or more cameras [8, 9]. The most frequent methods are several accelerometers and

gyroscopes coupled to different body locations [10 - 13]. Approaches combining visual and inertial sensors have also been considered [14]. Data processing is another critical component of all of these algorithms. According to some works, the performance is greatly influenced by the quality of the input characteristics concentrated on extracting the most valuable characteristics from a time series dataset [15]. The most typical method is to examine the signal in both the temporal and frequency domains. The active learning approach has been used in many machine-learning situations when identifying samples is time-consuming and costly. Speech recognition, information extraction, and unconstrained handwritten recognition are some examples of applications [16, 19, 20]. However, this method has yet to be applied to the human activity problem.

3. MAIN TECHNIQUES

K-Nearest Neighbor is a supervised learning technique that classifies the results of new instance queries using the majority of K-Nearest Neighbor categories. It is one of the most common activity detection algorithms. The KNN algorithm's goal is to store all existing objects and classify new ones using attributes and training samples. The KNN classifiers are memory-based and do not use any models to fit. As the prediction value of the new query instance, the KNN algorithm employed neighborhood classification. The KNN Algorithm is depicted in Fig. **1**. The k-nearest neighbor algorithm (K-NN) is a non-parametric (or distribution-free) method for classifying objects in the feature space based on the closest training samples. KNN is a memory-based learning method, where the function is just approximated locally, and the entire calculation is postponed until after categorization. The KNN algorithm is one of the most basic machine learning algorithms available: An object is classified by a majority vote of its neighbors, and the object is assigned to a class among its k nearest neighbors (k is a small positive integer). If k=1, the item is simply assigned to the class of its nearest neighbor. Each of the training examples is a vector with a labeled class in a multi-dimensional space. The feature vectors and class labels of the training samples are stored throughout the training (data pre-processing) phase of the algorithm. k is a user-defined constant in the classification phase of an activity recognition system, and an unlabeled vector is classed by it. Selecting the most common label from the k-training samples for continuous variables, Euclidean distance is typically applied as the distance metric

$$\text{Euclidean} \quad \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{1}$$

CHAPTER 11

Smart Water Bottle with Smart Technology

Dumpala Shanthi[1,*]

[1] *Maturi Venkata Subba Rao Engineering College, India*

Abstract: Working people have forgotten the importance of water intake. Smart Water Bottle will blow an alarm if it is being used by homemakers or people at home. This device can also be configured to their smartphones, which would send a message or alarm to their smartphones, making the user remember the intake of water by keeping some time interval. If configured to their mobiles, it would send a record of the amount of water taken by the user. The latest emerging technology like IoT, Android and many others will definitely be useful to the user.

Keywords: Smart Water Bottle, IoT, AI & ML, Arduino.

1. INTRODUCTION

The Internet of Things (IoT) is expected to grow exponentially to 50 billion devices by 2020 [1]. Smart homes, transportation, healthcare and industrial automation are new uses. Smart objects in our environment can provide context awareness that is often lacking in activity monitoring [2, 3]. Smart Stuff, or smart objects with physiological and activity sensors, used for new non-invasive physiological monitoring options. Smartphones and smart watches make it easy to connect SmartStuff to wearable sensor networks [4, 5].

New smartwatches like the Basis Peak and Apple Watch continuously measure physiological parameters like heart rate, GSR, and temperature. Many ubiquitous health monitoring applications require smartwatches to receive messages and notifications. Smart sensors and smart environments enable the quantified self (QS) [6, 7].

The QS community tracks physiological, behavioral, and environmental data. New sensors and systems enable easy record collection and database integration, allowing for data mining and new insights. Standard toolsets like Apple Health

[*] **Corresponding author Dumpala Shanthi:** Maturi Venkata Subba Rao Engineering College, India;
E-mail: Dshanthi01@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

Book, Google Fit, Samsung S.A.M.I, and Microsoft Health vault support this trend.

Big Data analytics can support personalized health monitoring and interventions [3, 7]. One of the most important factors for health and well-being is proper hydration.

60% of our body, 75% of our brain, and 83% of our blood contain water. An intelligent hydration monitoring and management platform can provide automatic, accurate, and reliable monitoring of fluid consumption and provide configurable advice for optimum personalized hydration management.

A smart water bottle can help a variety of people, from athletes to dieters to the elderly, who suffer from dehydration in group homes or alone. While most users who seek to improve their health do not drink enough water, some medical applications do. Patients with kidney disease or congestive heart failure (CHF) should limit their liquid intake while still drinking regularly throughout the day.

Besides getting notifications and reminders at the optimal times to drink, Current and historical intake, as well as progress towards personal hydration targets, can all be monitored by means of a unified hydration management platform. The original smart water bottle was called the Hydra Coach. There is a hydration indicator on the bottle, but it isn't wireless. Bluetooth wireless connections and specialised smartphone and smartwatch applications are featured on a number of recently released bottle designs [8]. This study demonstrates the functionality of a smart water bottle equipped with sensors.

2. EMBEDDED SYSTEM

Embedded systems are not intended to be general-purpose computers. Others may have low or no performance requirements, allowing the system hardware to be simplified and costs reduced [9].

An embedded system is often physically integrated into the device it controls. Embedded software is stored in read-only memory or flash convector chips rather than a disc drive. It frequently runs on computers with a small keyboard, screen, and memory [10, 11].

Embedded systems are controlled by microcontrollers or DSPs (DSP). The main characteristic is being dedicated to a specific task, which may necessitate powerful processors. For example, air traffic control systems use mainframe computers and dedicated regional and national networks to connect airports and radar sites. (Each radar probably has its own embedded system).

As embedded systems are task-specific, designers can optimize them to reduce product size and cost while increasing reliability and performance. Some embedded systems are mass-produced, saving money.

3. ARDUINO NANO

This breadboard is used to connect a few connectivities shown below in Table **1**.

Table 1. Arduino Nano Specification [9].

Arduino Nano	Specifications
Flash Memory	32 KB, of which 2 KB used by the Bootloader
Input Voltage	(7-12) Volts
Microcontroller	ATmega328P
Operating Voltage	5 Volts
PCB Size	18 x 45 mm
Power Consumption	19 milliAmps
PWM Output	6
SRAM	2KB
Weight	7gms

4. PIN DIAGRAM

4.1. Serial Communication

It can communicate with a computer, another Arduino, or other microcontrollers. The ATmega328 supports UART TTL (5V) serial communication on digital pins 0 (RX) and 1 (TX) (TX). An ATmega16U2 on the board sends serial data over USB and appears as a virtual com port to computer software. There is no need for an external driver with the 16U2 firmware. However, Windows requires an.inf file. The Arduino software includes a serial monitor for sending and receiving simple text data. The board's RX and TX LEDs will flash when data is sent over the USB-to-serial chip and to the computer (but not for serial communication on pins 0 and 1) in (Fig. **1**).

CHAPTER 12

Real World Applications of Machine Learning in Health Care

Kari Narmada[1,*], **Sanjay Kumar Singh**[1] and **Dumpala Shanthi**[2]

[1] *Amity University Rajasthan, Jaipur, Rajasthan, India*

[2] *Maturi Venkata Subba Rao Engineering College, Hyderabad, India*

Abstract: Machine learning (ML), a subset of artificial intelligence, is used to construct algorithms for monitoring, diagnosing, forecasting, and predicting clinical results. Health is a major concern for human beings. The current success in ML is due to deep learning (DL), using huge artificial neural networks. In the past, machine learning has demonstrated its usefulness and skills in detecting cancer. It is one of the most feasible solutions for top healthcare pioneers to detect anomalies. When healthcare companies succeed in using predictive models, they face challenges in demonstrating their value and gaining trust across the company. Recently, established standards for reporting machine learning-based clinical research will aid in connecting the clinical and computer science communities and realizing the full potential of machine learning techniques. The researchers have many objectives in the design of machine Learning Algorithms for different applications. Many papers discussed how machine learning algorithms are involved in health monitoring which will be updated so that patients, doctors, or any individuals can view the information. The main goal of this paper is to discuss basic types of Machine Learning and the challenges faced by Artificial intelligence (AI) in health care. The possible risks in clinical research give practical information on how to accurately and effectively analyze performance and avoid frequent pitfalls, particularly when dealing with applications for health and wellness contexts.

Keywords: Machine Learning , Health care , Applications , Challenges in AI .

1. INTRODUCTION

Nowadays, technology provides more rapid analysis to offer a proper treatment plan for human lives and health. Artificial Intelligence and Machine Learning are currently ruling the world. We know how significant advancements in Machine Learning Algorithms can benefit the healthcare industry in the medical field. IoT

[*] **Corresponding author Kari Narmada:** Amity University Rajasthan, Jaipur, Rajasthan, India;
E-mail: narmadakari@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

(Internet of Things) is the most common technology is Internet with things, providing excellent advantages in real-time applications.

In healthcare technology, machine learning algorithms are better at detecting malignant tumors, and it provides self-learning neural networks that improve treatment quality by analyzing data regarding the patient's health, such as numerous tests, X-rays, and screenings are all available. Deep learning algorithms are increasingly being utilized to detect cancer cells and improve clinical analyses. In general, Machine Learning can help with the treatment process in a variety of ways, such as automated messaging warnings [1].

A voice-controlled healthcare assistant functions as a computer-generated nurse for patients, providing information on a variety of ailments, health issues, and medications. Data engineers are developing solutions for overall health monitoring and assisting in treating or preventing disease.

The total number of cancer-diagnosed persons in the last two decades has almost doubled from a record 10 million in 2000 to 19.3 million in 2020. In today's world, one out of every five people will have cancer at any point in their lives. It is estimated that, in the coming years, there will be a much greater rise in the number of people living with cancer, and by 2040, it will be almost 50 percent over 2020. Cancer is the world's leading cause of mortality; almost 10 million deaths occurred in 2020. The WHO will hold the first consultations on World Cancer Day to develop a new global breast cancer plan starting in 2021 [2, 3].

Maintenance of good health with the present day-to-day life activities is not easy. So many humans does not have proper diet and exercise, leading to many health issues. Many human beings neglect their health and give priority to their daily lifestyle. This gives a clear chance for the diseases to attack easily. All over the world, around 8.9% of the population is suffering from diabetes in India and 9.3% worldwide. Similarly, the blood pressure patients are 13.8% in India and 19.3% worldwide. These values are exponentially increasing in the past decade, giving future generations bad results. Diagnosis done at the proper time will lead to control of these values. As an engineer, we can diagnose the patient, but if we create a channel between the patient and doctor, it helps the patients control their problems. Diseases are generally classified into two types [4].

- Acute diseases
- Chronic diseases

Nowadays, technology has been improved drastically; wearable systems and gadgets for health monitoring are manufactured by various readily available

companies. Modern simulated tools like IoT, machine learning, artificial intelligence, and data management are combined, and some superactive machines can be designed with these tools, which helps doctors to treat patients easily. A patient also easily interacts with the doctors without any disturbances [5]. The training of the data is done by using the data sets. Later testing is done. Testing always determines the non-regularities of the levels. Statistics is applied to the data set over the cloud, and analysis is done; finally, the accuracy is calculated. Acute conditions aren't too bad but last for life, while chronic conditions need surgery. Here are some examples of common acute disorders. Accuracy in the design of health monitoring systems is crucial to their success in curing many ailments, and in the modern world, everyone wants advanced technology; they want everything to be automated. Some businesses even use speech or voice recognition software to speed up processes [6].

2. LITERATURE REVIEW

The authors employed three machine learning algorithms to predict cancer cells. They extracted data from trials at the National Cancer Institute of Tehran using the ICBC dataset. On the ICBC dataset, they used three machine learning methods (ANN, D-Tree, and SVM).

Suji, R. J. *et al*. propose a system that combines CFRSFS, K-Mean clustering, and LS-SVM. CFRSFS is a feature selection method suggested by the author. To continue feature processing, CFRSFS selects just eight features and assigns them to K-Mean clustering. Their approach has a 99.54 percent success rate [7].

Rana, M *et al*., the authors used various Machine Learning algorithms based on the dataset. They obtained experimental results using the WDBC dataset. They assessed the accuracy of algorithms to be 68 percent for SVM-linear, 68 percent for logic regression, 72 percent for KNN, and 67 percent for Naive Byes [8].

Singh, S. *et al*. employed Neural Networks to identify breast cancer. They used 1808 cancer cases. For testing, they used 387 instances. They assessed the accuracy of the Neural Network method to be 95.80 percent.

Shattuck *et al*. updated the EM algorithm, and they made sure that the advanced EM was used for all the images. Wells *et al*. proposed an important technique called Bayesian tissue classification. They also designed an EM algorithm [9].

Davatzikos *et al*. discussed how the boundaries are extracted if it's in snake curves. The moving curves are very difficult to operate. They used a predefined algorithm to obtain the accuracy [10]. Cohen *et al*. proposed an excellent

CHAPTER 13

Investigating and Identifying Fraudulent Behaviors of Medical Claims Data Using Machine Learning Algorithms

Jyothi P. Naga[1,*], K.V.S.N. Rama Rao[2], L. Rajya[3] and **S. Suresh[4]**

[1] *Department of CSE, GITAM Institute of Technology, GITAM, Visakhapatnam, A.P., India*

[2] *Department of CSE, K L Educational Foundation, Hyderabad, A.P., India*

[3] *Department of Computer Science and Engineering, JNTUK UCEV, Vizianagaram, A.P, India*

[4] *Anil Neerukonda Institute of Technology & Sciences, Sangivalasa, Bheemunipatnam, Vishakapatnam, Andhra Pradesh, India*

Abstract: Healthcare is essential in pandemic times, but it is crucial for the well-being of daily life. Many countries allocate substantial funds towards providing high-quality healthcare services. As healthcare expenses escalate, policymakers and funders are increasingly focused on investigating the underlying factors driving the high costs of medical resources. A comprehensive analysis carried the required expenses towards identification, valuation, and measurement of resources utilized for the diagnosis process. The objective of the chapter is to provide how the data analysis is carried which helps to identify fraudulent behaviors. The generated model assists health management organizations in identifying suspicious behaviors toward claims. Healthcare fraud is a severe threat to global health results, and could lead to misuse, scarce resources, and negative impacts on healthcare access, infrastructures, and social determinants of health. Healthcare fraud is associated with increased healthcare costs in most of the leading countries. The proposed research work provides an estimation mechanism for utilizing health resources and their impacts on healthcare costs. This chapter proposes strategic ways of handling healthcare data to prevent future healthcare fraud, decrease healthcare expenditure, and adequately use resources to benefit the population. This chapter works on three primary datasets and a synthetic dataset aggregated from the primary datasets. The data preprocessing is carried out at different levels of the model, which truly enhances the data quality. The model is constructed at three levels; the first level analyzes datasets in which it extracts the primary features and provides constructive decisions and outcomes on the processing of data. Regressive analysis of the hierarchical grouping mechanism helps to know the detailed features that could affect healthcare and prevent resource misuse.

* **Corresponding author Jyothi P. Naga:** Department of CSE, GITAM Institute of Technology, GITAM, Visakhapatnam, A.P., India; npothaba@gitam.edu

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

Keywords: Healthcare, Machine Learning, Fraud detection, Statistical, Supervised, Decision.

1. INTRODUCTION

Many factors affect the health insurance systems in most countries, like increased population or fraud/deficient patients, fraudulent physicians, and providers, which automatically reduce the allotted funds by the government or companies. There is a significant increase in fraud, abuse, and waste of services, consequently straining finance. The scam includes suspected diagnosis proof, low insured willingness, recurrent claims, late filing of claims, inconsistent documents, insufficient damage level information, high payments, and poor reliability on the certificate. Each claim submitted must be approved by the medical professional to justify the chapter's weakness. It incorporates age, sex, name, date of birth, type of claim, identification number, resting period solicited, identification of medical expert, identification of employer and past and present income records.

Experts usually detect fraud within healthcare through auditors and investigators by searching numerous records manually and digging out suspicious or fraudulent behaviors. As the volume of medical data increases, it is tedious for experts to detect fraud. It may lead to weakness when compared to more automated data mining and machine learning methods. The global economic crime of 2018 by PwC reports that 49 percent of organizations suffer from fraud. According to the NHCAA, about 9% (47.9 billion dollars) of the United States, the annual healthcare expenditure is increased due to fraud and corruption. In Europe and Australia, the losses due to fraud and corruption correspond to 10% of the total expenses, according to the respective Healthcare Anti-Fraud and corruption institutions, EHFCN and PHA, Private Healthcare Australia. Every year this count increases to billion–dollars in business. The estimation of Insurance Fraud was nearly $80 to $100 billion in 2014, equating to the cost per insured household between $400 and $1000 in the United States [1, 2]. According to the Federal Bureau of Investigation, in 2013 cost of all insurance fraud was about $40 billion, with about 20 percent being in the property-casualty sector [3 - 5]. European Insurance Anti-Fraud Guide has estimated the cost of fraud cannot be less than €8 billion or approximately 2 percent of the total annual premium income from all classes of the European insurance market in 1996. In Australia, 10 percent of the total amount is paid for fraudulent claims each year which is about AUS$1.4 billion [6].

In contrast, in Latin America and the Caribbean, the fraud cost is estimated at 19 percent to 35 percent of the annual revenue of the insurance industry. Only due to claim fraud/policyholder fraud, South Africa had lost 100 million rands in 2010

[7]. In Kenya, Kuria, and Morongo, it is estimated that 40 percent are fraudulent claims, and it is between 10 percent to 30 percent in Nigeria [8]. Health Insurance Portability and Accountability Act (HIPAA) is a US regulatory system to punish the criminals involved in insurance fraud based on the extent of the fraud. The possibilities of fraud are co-evolving with technology. Fraud analytics is a probe towards digital technology across industries. Fraud prevention and detection have to chapter simultaneously to stop the occurrence of fraud and handle it in the best possible way. In practice, detecting fraud starts with the prevention of false alarms, like identifying the tradeoff between the costs of analyzing a fraudulent case and saving by investigating it by taking it into consideration. This chapter highlights opportunities and opportunities and challenges researchers by handling multiple attributes that assess the medical service cost for different procedures. Medical service costs will somehow decrease the high costs of resources, increase the scrutiny by payers, and reflect the payment to payers' actual negotiated price. Therefore, resource utilization will be more effective and may be reduced to low-end procedures.

To summarize, some factors are influencing the cost of International Health Insurance in a majority of countries like:

- There is an increased demand for the high-quality healthcare
- The increased cost of healthcare
- Increased regulations and regulatory systems
- Continuing challenges related to fraud

The issues of bribery and fraudulent claims remain the same as the most significant challenges to healthcare sectors. Hence there is a need for automated models to meet the above requirements.

2. ROLE OF MACHINE LEARNING ALGORITHMS

Manual Processing of health care data may lead to wrong diagnoses and analysis. The power of using electronic technology for recording patient information to improve the chapter flow is somehow remarkable. Using advanced analytics, it can provide better information on patient care to doctors. There are several successful cases that machine learning algorithms are able to identify cancer tumors, skin cancer and diabetic retinopathy. Machine learning helps for quicker decision-making. AI and ML can automate the routine task in the forthcoming years. This could open more number of possibilities for human to take over more sophisticated tasks. The various applications of ML in the medical domain like robot-assisted surgery, virtual nursing, administrative assistants, fraud detection,

CHAPTER 14

Security Threats and Detection Mechanisms in Machine Learning

K. Madhuri[1,*]

[1] *Anurag University, Hyderabad, India*

Abstract: Machine Learning refers to computer programming or data science that can learn from data. For example, the performance of T is increased with experience E if a computer program is learned from E in some tasks T, and performance is measured as P. The learning of computers is allowed without explicit programming in Machine Learning, which is a field of research. Machine learning can be used in various applications, such as banking, travel and tourism, healthcare, marketing, insurance, and human resources. Machine learning is a powerful tool for implementing security applications.

Keywords: Machine learning, AI, Security, threats, Classification, Regression, Fraud Detection, Spam Detection.

1. INTRODUCTION

ML is a subcategory of AI that uses algorithms to achieve automatic performance without prior programming. For predicting new output values, historical data will be used in machine learning algorithms.

Machine learning includes different categories, such as:

1. Reinforcement machine learning

2. Semi-supervised machine learning

3. Unsupervised machine learning

4. Supervised machine learning

* **Corresponding author K. Madhuri:** Anurag University, Hyderabad, India; E-mail: madhuris97@gmail.com

Dumpala Shanthi, B. Madhuravani, & Ashwani Kumar (Eds.)

For training the algorithms, the labelled data sets are used in supervised learning for data classification or determining the correct outcomes. Supervised learning is one of the types of machine learning that uses the 'labelled' training data for training, and the results are predicted using this data. The marked data describes marking some input data with the correct output. The training data provided to the machine considers a supervisor in supervised learning, and the machine learns about how to predict the output accurately. It also applies the concept of students learning under the guidance of a teacher. The process of providing inputs and outputs correctly for machine learning models is supervised learning. It has the objective of determining a mapping function to correlate the input variable (x) and the output variable (y). Supervised learning involves real-time applications, such as spam filtering, fraud detection, image classification and risk assessment. Based on the labelled data sets, the training of the model is carried out in supervised learning, and each type of data is learned by the model. The model will test using the test data after completing the training process, and the results will be predicted [1].

1.1. Supervised Learning Example

For example (Fig. **1**), if we have a data set that includes different types of shapes like polygons, triangles, rectangles, and squares. For each shape, the model is trained in the first step. The shape will mark as a square when a given shape includes four sides, and they all are equal.

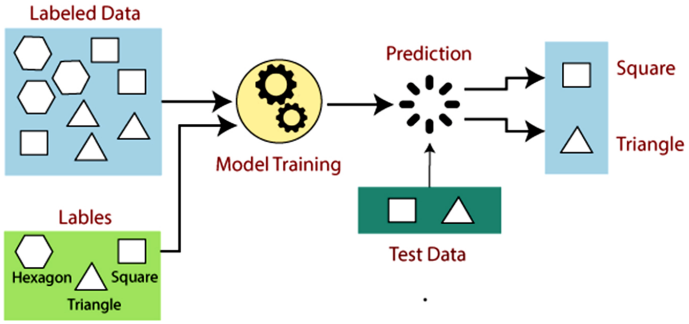


Fig. (1). Supervised Learning Example.

If a shape includes three sides, it will mark as a triangle. Hexagon will be marked when a shape includes six equal sides. For testing the model, the test set is used after training. The model has the task of recognizing the shape. The training of a machine has been performed on different shapes. Based on the edges, the shapes will be sorted out if a new shape is found and the result is estimated.

1.1.1. The Steps Involved in Supervised Learning

Primarily, the training data set type needs to determine and collect the labelled training data, which is categorized into a verification data set, test data set, and training data set. Sufficient knowledge should have been included for computing the input characteristics of the training data set that leads to the accurate prediction of output. The appropriate algorithms estimate for the model, such as decision trees, support vector machines, *etc*. The algorithm runs on the training data set. The validation set is required as a control parameter, which is a subset of the training data set. A test set is provided to evaluate the model's accuracy. The model is accurate when the model predicts the correct output [2].

Similar to the learning of things by students in the presence of a teacher, supervision is required for training the model in supervised learning, which can be utilized for two different problems, such as regression and classification, as shown in Fig. (**2**).

1.1.2. Supervised ML Algorithm Types

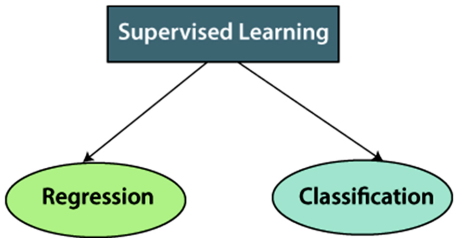


Fig. (2). Supervised ML algorithm types.

1.1.3. Regression

The regression algorithm is used when a relationship exists between input and output variables. It is used to predict continuous variables like market trends, weather forecasts, *etc*. Some popular regression algorithms are included in supervised learning:

o Linear regression

o Regression tree

o Nonlinear regression

o Bayesian linear regression

SUBJECT INDEX

A

Activity 195, 197, 200, 204
  recognition process 200
  recognition system 195, 197
  sensors 204
Adaptive 45, 71, 268
  neuro-fuzzy inference system 45, 71
  spam filtering technology 268
AI-based 106, 226
  prediction system 106
  technology 226
Algorithm(s) 4, 8, 9, 10, 12, 13, 23, 24, 25, 26, 29, 46, 53, 60, 61, 81, 128, 134, 176, 178, 179, 196, 222, 223, 227, 263
  complicated 176
  constant 29
  data-mining 81
  generic 46
  hereditary 60
  transparency 227
Alzheimer's disease 84
Arabidopsis thaliana and prediction of nutrient deficiency 100
Arduino software 206, 207
Artificial 1, 3, 4, 45, 60, 61, 62, 64, 66, 67, 70, 71, 105, 134, 140, 268, 271
  immune system 268
  intelligence-based techniques 271
  intelligence methods 105
  neural networks (ANNs) 1, 3, 4, 45, 60, 61, 62, 64, 66, 67, 70, 71, 134, 140
Automated 105, 106, 107
  monitoring system 106
  plant health monitoring system 105, 106, 107
  selective breeding system 105, 106
Automatic 82, 212
  clustering and under sampling (ACUS) 82
  gain control (AGC) 212

B

Bacterial leaf blight (BLB) 104
Bayes theorem 61, 268
Bayesian 13, 25, 45, 57, 69, 71, 137
  networks 13, 45, 69, 71
  prediction technique 45, 57, 71
  technique 25
  technology 137
  theorem 25
Benchmark datasets 83
Big data analytics (BDA) 1, 15, 16, 205, 236
Biotic stresses 105
Blood transfusion 228
Brain 21, 51, 133, 138, 205
  artificial 138
  images 51, 133
Breast cancer 49, 50, 51, 222
  detection 51
  diagnosis 51

C

Cancer 49, 50, 90, 220, 221, 225, 233, 262
  detection 90, 225, 262
  skin 225, 233
  tumors 233
Characteristics, stress resistance 106
Chronic diseases 221
Cloud 1, 14, 16
  computing 1, 14, 16
  security 14
Cloud-based 14
  approach 14
  solution 14
Clustering 10, 46, 56, 57, 82, 224
  algorithms 10, 57
  approaches 46
  method 224
  methodologies 82
  techniques 56
Communication 215, 265
  bandwidth 265

devices 215
Comparative aggregate data analysis 243
Computer 3, 13, 14, 19, 67, 76, 108, 206, 220, 226, 255, 265
algorithms 19
networks 13, 14, 265
programs 76, 255
science communities 220
software 206
system 3
vision 67, 108, 226
Congestive heart failure (CHF) 205
Convolutional neural networks (CNN) 45, 46, 50, 51, 70, 88, 113, 114, 115, 118, 140
Crop 11, 97, 106
diseases 97
health 106
management 11
Cryptography 224
Cyber security attacks 14

D

Damage 225
nerve 225
DarkNet database 122
Data 89, 197, 240, 247, 269
preprocessing techniques 89
processing 197, 240, 269
transformation 247
Data mining 13, 16, 20, 45, 56, 57, 61, 65, 66, 68, 71, 79, 81, 90, 162, 204, 232, 239
algorithms 66, 79
and image processing 162
automated 232
methods 66, 90
techniques 45, 56, 65, 66, 68, 71, 239
traditional 81
Datasets 77, 81, 131, 239, 240, 251
aggregated 239, 240, 251
aircraft 77
cardiovascular 81
stored 131
Decryption algorithms 223
Deep 104, 137, 138, 140
architecture 140
reinforcement learning 137, 138, 140
transfer learning 104
Deep learning (DL) 2, 10, 14, 87, 88, 97, 100, 101, 107, 108, 114, 137, 138, 140, 220, 221
based machine vision methods 114
techniques 10
Detection 15, 112, 114, 116, 233, 234, 235, 252, 257, 259, 261, 265, 266, 267, 269, 270, 271
abuse 270
anomaly 235, 270, 271
automatic lesion 15
fraudulent claim 252
techniques 235
Devices 5, 7, 14, 29, 33, 204, 205, 208, 209, 210, 211, 212, 213, 215, 216, 223, 225
automatic surveillance 5
intrinsic photoelectric 208
multiple 212, 215
smart medical 225
storage 7
Diabetic 225, 233
renal disease 225
retinopathy 233
Diagnose 51
breast cancer 51
skin cancer 51
Disease(s) 49, 53, 75, 76, 87, 88, 89, 98, 100, 102, 104, 106, 107, 108, 221, 234
infected 102
infectious 87
malaria 89
prediction, effective cardiovascular 53
DL techniques 97, 108
DM techniques 237

E

Electronic fraud detection, (EFD) 235
ENN 144, 152
based opposition behavior learning 144
classification 152
Ensemble learning techniques 86

F

Factors, resistance-related transcription 106
Flood prediction 63, 64
accuracy 64
methods 64

Forecasting rainfall 62, 65, 66
Fraud 234, 251, 271
  digital 271
  in health insurance 234
  medical 251
  policies 251
Fraud detection 75, 232, 233, 235, 237, 238, 255, 256, 258, 270, 271, 273
  credit card 75
  electronic 235
  methods for 235
Fraudulent transactions 270

G

Gaussian mixture modelling 116
Genomic(s) 98, 99, 100, 108
  selection (GS) 99, 100
  technologies 98, 108
Gesture datasets 87
Google neural machine translation 6, 35
Gradient boosting machines (GBM) 103
Granular support vector machine 81
Grasshopper optimization 157

H

Hardware engineers 223
Health insurance systems 232
Healthcare 27, 29, 220, 221, 226, 229
  data analysis 226
  industry 27, 29, 220
  machine learning 229
  technology 221
Hidden conditional random field (HCRF) 87

I

IBM's Watson genomics 225
ICBC dataset 222
Identifying medical fraud 238
Image(s) 51, 107, 129, 132, 135, 140, 226
  medical 129, 132, 135, 140, 226
  microscopic biopsy 51
  segmentation methods 107
Image processing 31, 50, 107, 108, 162
  algorithms 31
  methods 108
  process 107
  techniques 50
Imbalanced malaria datasets 85
Index of balanced accuracy (IBA) 83
Inflammatory cytokines 89
Information 97, 98, 100, 236
  genetic 98
  genomic 97, 100
  hospital 236
Interactive voice response system (IVRS) 2
Internet of Things (IoT) 69, 204, 220, 221, 222, 224

L

Language properties 182
Latent dirichlet allocation (LDA) 137, 180
Learning 61, 196, 225, 237
  algorithms 196
  approaches 225
  methods 61, 237
LIFELAB dataset 79
Light 101, 208, 209, 210
  dependent resistor (LDR) 208, 209, 210
  interception 101
Linear 12, 19, 24, 40, 45, 56, 60, 70, 71, 137, 257, 261, 262
  discernment analysis 137
  regression 12, 19, 24, 40, 45, 56, 60, 70, 71, 257, 261, 262
Liquid crystal display 213
Logistic 10, 11, 45, 65, 66, 70, 71, 80, 237, 238, 261, 262, 263
  regression (LR) 10, 11, 45, 65, 66, 70, 71, 80, 237, 238, 261, 262, 263
  regression algorithms 263
  regression approaches 66
Long short-term memory (LSTM) 3

M

Machine, idle 14
Machine learning 9, 21, 45, 46, 54, 58, 61, 79, 90, 196, 225
  algorithms 9, 54, 79, 196
  applications 225
  approaches 58, 61
  reinforcement 255
  systems 21, 45, 46
  traditional 90

Mammography 50
Markov 46, 70, 87
  chain Monte Carlo sampling technique 87
  decision process 70
  models, hidden 46
Matching phonetic transcriptions 3
MATLAB 13, 56
  classification toolbox 13
Matthew's correlation coefficient 237
Memory 162, 212, 267
  analysis 162
  based machine learning techniques 267
  cells 212
Meta-search engine 147
Microprocessor control 211
Multi 83, 102
  dimensional scaling (MDS) 83
  linear regression (MLR) 102
Multimodal fusion process 158
Multivariate alteration detection (MAD) 114, 116

N

Naïve Bayes 268, 270
  algorithm 268
  spam detector 270
Natural 2, 6, 11, 33, 63, 65, 140, 171, 273
  catastrophes 65
  disasters 63
  language processing (NLP) 2, 6, 11, 33, 140, 171, 273
Naval air training 87
  and operating procedures standardization (NATOPS) 87
Nave-Bayesian classification 51
Network 3, 146, 213, 216
  launching wireless community 216
  microphone 213
  multilayered neural 3
  social 146
Neural network techniques 54
Nutrient deficiency 100

O

Oil content of sesame (OCS) 102
Olympic Games 187
Opposition 156
  based learning 156
  behavior learning (OBL) 156
Orthogonal complement subspace projection (OCSP) 84

P

Particle swarm optimisation (PSO) 45, 71
Phenotyping 101, 103
Plant 98, 101, 103, 105, 108
  breeding 103, 108
  disease discovery 105
  Genomics in agriculture 98
  health monitoring processes 108
  inheritance 98
  phenotyping 101
  rice 101
Power consumption 206, 217
Prevention, aid epidemic disease 227
Principal component analysis (PCA) 10, 65, 71, 135

R

Radial basis function (RBF) 82, 83, 103
  network 82
Rainwater harvester 62
Random forest (RF) 3, 19, 25, 70, 71, 86, 87, 89, 90, 101, 103, 104, 236, 237, 249, 250, 251
  algorithm 3, 101
  techniques 89
Ranking techniques 159
Recurrent neural network (RNN) 3, 4, 45, 67, 70, 71, 140, 270
Recursive feature elimination (RFE) 101
Regression 23, 24, 25, 62, 65, 69, 71, 80, 115, 116, 137, 138, 255, 257, 258, 259, 260, 261, 263
  algorithm 24, 257
  analysis 259, 260, 261
  problems 115, 116, 258
  process 138
  task 23
  technique 25, 65, 261
  tree 80, 257

S

Satellite data 227
Scale-invariant feature transform (SIFT) 38, 114
Search engine optimisation 16
Sepsis 228
SIFT techniques 116
Signals 61, 87, 197, 200, 211, 212
  aircraft handling 87
  microphone 211
  rotation 200
Simple 266, 268
  mail transfer protocol (SMTP) 266
  processing components 268
SMOTE 83, 89
  bagging (SBAG) 89
  borderline 83
  technique 83, 89
Social media 170, 270
  interactions 170
  platforms 270
Software 20, 51, 57, 58, 59, 60, 86, 88, 135, 144, 162, 222, 226
  bugs 58, 59
  components 60
  defect prediction datasets 86
  defects 60, 86
  engineering 162
  healthcare analytics 226
  logical modeling 88
  reliability 58
  tool 57
  voice recognition 222
Soybean farming 99
Spam 264, 268
  classification methods 268
  detection algorithm 264
Spam filtering 47, 256, 258, 266, 267, 268
  techniques 266, 267
Speech recognition 5, 15, 140, 197
Spherical topology 134, 223
Staple crop production 99
Stress 106, 196
Support Vector Machine (SVM) 13, 14, 19, 25, 45, 56, 57, 64, 66, 71, 80, 83, 102, 103, 114
SVM 56, 101
  classification method 101
  method 56
System 76, 98
  artificial 76
  automated 76
  ecological 98
System authentication 216

T

Techniques 11, 12, 13, 14, 16, 59, 61, 64, 65, 81, 83, 84, 87, 100, 112, 130, 131, 132, 135, 145, 156, 235, 267, 268
  bootstrapping 87
  data-mining 61
  indexing 145
  meta-evaluation 132
  over-sampling 81, 83, 84
Text mining 90, 165, 166, 273
Touch sensor 217
Traditional 7, 20, 40, 98, 114
  based machine vision methods 114
  breeding techniques 98
  programming 7, 20, 40
Transactions 262, 270, 271, 272
  financial 270
  fraudulent monetary 272
Transcription factors (TFs) 106
Transfer learning methods 107
Tumor 221, 263
  malignant 221
  size 263

V

Vectors 11, 121, 197, 267
  multi-dimensional space 267
Vehicle 113, 116
  density 116
  direction 116
  tracking 113
Vehicle detection 113, 114, 116
  vision-based 114
Virtual nursing 233

W

Web mining 144, 159, 160
Website customization re-ranking technique 145
WIFI technology 215

Wireless local area networks (WLAN) 215
Wisconsin prognostic breast cancer (WPBC) 51
Word 17, 165, 186, 187, 189
  cloud explorer 165, 186, 187
  spotting approach 17
  spotting text analytics 165, 189

Dumpala Shanthi

Prof. Dumpala Shanthi has 15 years of experience in teaching and training. She is a part of many prestigious events like Nan Mudhalvan, Nalayiah Thiran, Productkraft Skills and PMVY programs as a freelancer. She completed Postdoc from California Public University, USA, completed PGP diploma course on AI & ML in NITW. She received Ph.D. in CSE from JNTU Hyderabad in 2020. She has good academics & research experience in various areas of computer science. Her area of research includes software engineering, artificial intelligence, machine learning, biometrics and machine learning. She has published a few national & international patents successfully and received the grant certificate. She has published a number of research papers in various reputed national and international journals (including IEEE, Springer, Inderscience etc.) and proceedings. She has been awarded the "Best Paper Award" in 2019, "Young Scientist Award" by DK International Research Foundation Faculty Awards 2019, etc. She has published a few books at national and international levels. She is a member of various professional bodies such as, CSI, IEEE and IAENG. She is an editorial board member of "American Journal of Software Engineering and Applications", IJERT, Micro Processors and Micro Controllers (SCI Journal).

B. Madhuravani

Prof. B. Madhuravani has over 12 years of teaching experience and presently working as professor in the Department of Computer Science & Engineering, MLR Institute of Technology, Dundigal, Hyderabad, Telangana, India. She received Ph.D. in faculty of CSE from JNTU Hyderabad in 2019. She is the author or coauthor of more than 40 papers in the international refereed journals and more than 10 conference contributions. She has published 6 Indian and 1 international patents. She has been granted 1 patent and 1 copyright, and has published 2 books. She has been working since 2012 in research projects related to information security and successfully completed the DST project as principal investigator under Women Scientist Scheme (WOS-A). She has received several important recognitions to her Professional and Research career, including the "Excellence Teaching in Higher Education Award" by DK International Research Foundation Faculty Awards 2019 and "Outstanding Research Faculty Award" by Amaravathi Research Academy Faculty Excellence Awards 2020. Her area of interest includes security, machine learning, artificial intelligence, data science, big data and wireless sensor networks. She is a member of various professional bodies, such as ISTE, CSI, IEEE and IAENG.

Ashwani Kumar

Prof. Ashwani Kumar is an associate professor in the Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, India. He received Ph.D. in Computer Science and Engineering, from Jaypee University of Information Technology, H.P in 2017. He received M.Tech (CSE) from Jaypee University of Information Technology, H.P in 2011. He has done B.Tech. from Uttar Pradesh Technical University (U.P.T.U) Lucknow, U.P in 2009. He has more than 12 years of teaching and research experience. He published more than 50 research papers in international reputed journals indexed in SCI/SCIE Web of Science & SCOPUS and top rank conferences as authored or co-authored. His research findings are published in Wiley, Springer Journal, Procedia Computer Science Elsevier, JST Universiti Putra Malaysia, IJCAC IGI Global USA, IEEE Xplore. He is serving as the reviewer of many SCI Journals- IET Information IET Digital Library, Multimedia Tools and Application (MTAP) Springer, National Academy Science Letters Springer, Journal of Intelligent and Fuzzy Systems (JIFS) IOS Press. Currently, he is editorial board member of many reputed Journals-IJDCF, IJISP, IJEUCD, IGI Global Journal USA indexed in WOS/SCOPUS. He has been the guest editor for International Journal Recent Patent of Engineering, Bentham Publication indexed in SCOPUS. He is the editor of the book series: Advancements in Security and Privacy Initiatives for Multimedia Images, IGI Global Journal USA. He received Bentham Brand Ambassador Award for the year 2019-2020. He served as session chair in several international conferences-ICIIP-2019, MISS-2020, IC4S-2020, PGDC-2020. He has filed an Australian patent in the year 2020. His research area includes image security, watermarking techniques, multimedia application, security & privacy, cryptography and network security. He is also a lifetime member of professional societies like ISTE, IACSIT, IAENG and SDIWC.