Lecture Notes in Networks and Systems 648

Ajith Abraham ·
Thomas Hanne · Niketa Gandhi ·
Pooja Manghirmalani Mishra ·
Anu Bajaj · Patrick Siarry *Editors*

Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022)

Springer

Lecture Notes in Networks and Systems 648

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

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

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Ajith Abraham · Thomas Hanne · Niketa Gandhi ·
Pooja Manghirmalani Mishra · Anu Bajaj ·
Patrick Siarry
Editors

Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022)

Springer

Editors
Ajith Abraham
Faculty of Computing and Data Science
FLAME University
Pune, Maharashtra, India

Scientific Network for Innovation
and Research Excellence
Machine Intelligence Research Labs
Auburn, AL, USA

Niketa Gandhi
Scientific Network for Innovation
and Research Excellence
Machine Intelligence Research Labs
Auburn, AL, USA

Anu Bajaj
Computer Science and Engineering
Department
Thapar Institute of Engineering
and Technology
Patiala, Punjab, India

Thomas Hanne
University of Applied Sciences and Arts
Northwestern Switzerland
Olten, Switzerland

Pooja Manghirmalani Mishra
Scientific Network for Innovation
and Research Excellence
Machine Intelligence Research Labs
Mala, Kerala, India

Patrick Siarry
Université Paris-Est Créteil
Créteil, France

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-27523-4 ISBN 978-3-031-27524-1 (eBook)
https://doi.org/10.1007/978-3-031-27524-1

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

Preface

Welcome to the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022) and the 14th World Congress on Nature and Biologically Inspired Computing (NaBIC 2022) held in the World Wide Web during December 14–16, 2022. Due to the ongoing pandemic situation, both events were held online.

SoCPaR 2022 is organized to bring together worldwide leading researchers and practitioners interested in advancing the state of the art in soft computing and pattern recognition, for exchanging knowledge that encompasses a broad range of disciplines among various distinct communities. The themes for this conference are thus focused on "Innovating and Inspiring Soft Computing and Intelligent Pattern Recognition". SoCPaR 2022 received submissions from 20 countries, and each paper was reviewed by at least five reviewers in a standard peer-review process. Based on the recommendation by five independent referees, finally 69 papers were presented during the conference (acceptance rate of 37%).

Nature and biologically inspired computing brings together international researchers, developers, practitioners, and users. NaBIC invited authors to submit their original and unpublished work that demonstrates current research in all areas of nature and biologically inspired computing, as well as industrial presentations, demonstrations, and tutorials. NaBIC 2022 received submissions from 15 countries, and each paper was reviewed by at least five reviewers in a standard peer-review process. Based on the recommendation by five independent referees, finally 19 papers will be presented during the conference (acceptance rate of 36%).

Many people have collaborated and worked hard to produce this year successful HIS–IAS conferences. First and foremost, we would like to thank all the authors for submitting their papers to the conference, for their presentations and discussions during the conference. Our thanks to program committee members and reviewers, who carried out the most difficult work by carefully evaluating the submitted papers. Our special thanks to the following plenary speakers, for their exciting plenary talks:

- Kaisa Miettinen, University of Jyvaskyla, Finland
- Joanna Kolodziej, NASK- National Research Institute, Poland
- Katherine Malan, University of South Africa, South Africa
- Maki Sakamoto, The University of Electro-Communications, Japan
- Catarina Silva, University of Coimbra, Portugal
- Kaspar Riesen, University of Bern, Switzerland
- Mário Antunes, Polytechnic Institute of Leiria, Portugal
- Yifei Pu, College of Computer Science, Sichuan University, China
- Patrik Christen, FHNW, Institute for Information Systems, Olten, Switzerland
- Patricia Melin, Tijuana Institute of Technology, Mexico

Our special thanks to the Springer Publication team for the wonderful support for the publication of these proceedings. We express our sincere thanks to the session chairs and organizing committee chairs for helping us to formulate a rich technical program. We express our sincere thanks to the organizing committee chairs for helping us to formulate a rich technical program. Enjoy reading the articles!

Ajith Abraham
Patrick Siarry
Thomas Hanne
General Chairs

Isabel Jesus
Nazar Zaki
Program Chairs

SoCPaR 2022 Organization

Patrons

Ashok Shettar (Vice Chancellor)	KLE Technological University, India
B. L. Desai (Executive Dean)	KLE Technological University, India

General Chairs

Ajith Abraham	Machine Intelligence Research Labs (MIR Labs), USA
Patrick Siarry	Université Paris-Est Créteil, France
Thomas Hanne	University of Applied Sciences and Arts Northwestern Switzerland, Switzerland

Program Chairs

Isabel Jesus	Institute of Engineering of Porto, Portugal
Nazar Zaki	United Arab Emirates University, UAE

Publication Chairs

Niketa Gandhi	Machine Intelligence Research Labs, USA
Kun Ma	University of Jinan, China

Special Session Chair

Gabriella Casalino	University of Bari, Italy

Publicity Chairs

Pooja Manghirmalani Mishra	University of Mumbai, India
Anu Bajaj	Machine Intelligence Research Labs, USA

Publicity Team

Peeyush Singhal	SIT - Pune, India
Aswathy S. U.	Jyothi Engineering College, India
Shreya Biswas	Jadavpur University, India

International Program Committee

Agostino Forestiero	CNR-ICAR, Italy
Ajaya Kumar A.	Gayatri Vidya Parishad College of Engineering, India
Alessandra Vitanza	University of Bari, Italy
Alexey Kornaev	Innopolis University, Russia
Alfonso Guarino	University of Foggia, Italy
Amit Kumar Mishra	DIT University, India
Anthony Gillioz	University of Bern, Switzerland
Anu Bajaj	Thapar Institute of Engineering and Technology, India
Arijit Ghosal	St. Thomas' College of Engineering and Technology, India
Ashwin Dobariya	Marwadi University, India
Aswani Cherukuri	VIT University, India
Aswathy S. U.	Marian Engineering College, India
B. Veera Jyothi	Chaitanya Bharathi Institute of Technology, India
Benjamin Fankhauser	University of Bern, Switzerland
Corina Masanti	University of Bern, Switzerland
Debdatta Kandar	North-Eastern Hill University, India
Deepika Koundal	University of Petroleum and Energy Studies, India
Durga Prasad B.	Nalla Narasimha Reddy Education Society's Group of Institutions, India
Elizabeth Goldbarg	Federal University of Rio Grande do Norte, Brazil
Emiliano del Gobbo	Università degli Studi di Foggia, Italy
Fariba Goodarzian	University of Seville, Spain
G. R. Anil	Vardhaman College of Engineering, India
Gabriella Casalino	Università degli studi di Bari "A. Moro", Italy
Gianluca Zaza	University of Bari "Aldo Moro", Italy
Giovanna Adele Naselli	Italian Institute of Technology (IIT), Italy
Giovanna Castellano	Università di Bari, Italy
Hare Ram Sah	Sage University, India
Hrushikesava Raju	KL University, India
Hudson Geovane de Medeiros	Federal University of Rio Grande do Norte, Brazil

Jagini Nagapadmaja	Vardhaman College of Engineering, India
José Everardo Bessa Maia	State University of Ceará - UECE, Brazil
Kantipudi MVV Prasad	Symbiosis Institute of Technology, India
Kaspar Riesen	University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
Kolla Bhanu Prakash	Koneru Lakshmaiah Education Foundation, India
Krishna Keerthi Chennam	Vasavi College of Engineering, India
Lakshmana Rao Kalabarige	GMR Institute of Technology, India
Lee Chang-Yong	Kongju National University, South Korea
Lokeshwari Vinya Viyyapu	Vardhaman College of Engineering, India
M. Ravi	CMR Institute of Technology, India
Madhuri M.	Chaitanya Bharathi Institute of Technology, India
Mahendra Kanojia	Sheth L.U.J and Sir M.V. College, India
Manisha Guduri	BIT Mesra, India
Mathias Fuchs	University of Bern, Switzerland
Mayssa Frikha	Sfax University, Tunisia
Mohamed Sirajudeen Yoosuf	VIT University, India
Murilo Machado	Universidade Federal de Santa Maria, Brazil
MVV Prasad Kantipudi	Symbiosis Institute of Technology, India
Oscar Castillo	Tijuana Institute of Technology, Mexico
Paolo Pagliuca	National Research Council, Italy
Pasquale Ardimento	University of Bari “Aldo Moro”, Italy
Pooja Manghirmalani Mishra	Machine Intelligence Research Labs, India
Prabu U.	Velagapudi Ramakrishna Siddhartha Engineering College, India
Radu-Emil Precup	Politehnica University of Timisoara, Romania
Rajanikanth Aluvalu	CBIT, India
Raman Dugyala	Chaitanya Bharathi Institute of Technology, India
Ramana Kadiyala	Chaitanya Bharathi Institute of Technology, India
Ramana Reddy B.	Chaitanya Bharathi Institute of Technology, India
Sangeeta Gupta	Chaitanya Bharathi Institute of Technology, India
Satyanarayana Murthy T.	Chaitanya Bharathi Institute of Technology, India
Senthilkumar Mohan	VIT University, India
Shobarani Salvadi	Chaitanya Bharathi Institute of Technology, India
Shubham Mahajan	Ajeenkya D Y Patil University, India
Shweta Paliwal	DIT University, India
Sidemar Fideles Cezario	Universidade Federal do Rio Grande do Norte, Brazil
Sindhu P. M.	Nagindas Khandwala College, India
Sravan Kumar	Sreyas Institute of Engineering and Technology, India
Srinivasa Routhu	NIT Kurukshetra, India

Suryanarayana G.	Vardhaman College of Engineering, India
Swapna Mudrakola	Matrusri Engineering College, India
Thatiana C. N. Souza	Federal University of Rio Grande do Norte, Brazil
Tushar Jaware	R C Patel Institute of Technology, India
Tzun-Pei Phong	National University of Kaohsiung, Taiwan
Uma Maheswari V.	KG Reddy College of Engineering and Technology, India

Contents

Soft Computing and Pattern Recognition

Multiple Vehicle Tracking Using Meanshift Algorithm and 8-point Connectivity

Prabha Nissimagoudar, Nihal Algur(✉), Nikhil Bonageri, Apeksha Chavan, Abhishek Koppa, and Nalini C. Iyer

KLE Technological University, Hubballi, India
{pcnissimagoudar,nalinic}@kletech.ac.in, nihalalgur9@gmail.com

Abstract. The rise in the number of accidents on the road due to human error is one of the major concerns of safety and has eventually paved way to the boom of autonomous vehicle industry. One of the most important features in automotive vehicles is the vehicle tracking which has led to the development of the automotive industry. Tracking of vehicles is one of the significant challenges faced in the automotive industry. Its main objective is to determine the exact path and direction of the vehicles. This project gives an elaborate view and understanding on why tracking in autonomous vehicles is an integral aspect of the industry. This project employs mean shift algorithm to construct a framework that is useful for single vehicle tracking. Carla simulator has been utilised for implementing pre image processing and also the distance of the obstacles from ego vehicle has been calculated. Further when mean shift algorithm was not helpful to track multiple obstacles, 8-point connectivity solved this issue. The successful implementation of multiple vehicle tracking has been done.

1 Introduction

According to the government statistics, vehicle accidents are the main reason for hospital admissions, fatalities and impairments in the nation. In India, traffic accidents claim the lives of over 1,50,000 people annually. The main contributors to the collision of vehicles are rash driving, poor traffic infrastructure and more. To overcome this problem, automation in the vehicles is necessary to improve safety as it takes out human intervention as much as possible and therefore reducing human error. Hence this makes transportation more safe and reduces road congestion, crashes and saves many lives.

An autonomous vehicle is a combination of different networking systems and sensors that assist the vehicle in driving on its own, based on its own intelligence. For an autonomous vehicle, tracking the approaching vehicles, objects or humans becomes an essential task to assist the car in making time-critical decisions. Tracking basically recognizes immobile as well as in motion vehicles in the surroundings. The inputs given to the vehicle are images through sensors. The process of tracking has been achieved with the use of multiple algorithms.

Some are fusion based approaches like [1, 2], in which LiDARs, Cameras (Monocular and stereo), and Radars in a wide range of configurations. Based on their sensor configuration and computational requirements, these approaches either Studies like [3]

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 3–13, 2023.
https://doi.org/10.1007/978-3-031-27524-1_1

and [4] Studies such as citeb3 and citeb4 rely only on stereo vision to perform car tracking. These sensors are restricted by the quality of the stereo pair and the field of vision, and so cannot track objects/vehicles in their entirety. Whereas studies like [5] and [6] explains the single camera strategy to analysing surrounding vehicle behaviour, but are severely limited and constrained by the camera's FOV and localization capabilities. A major problem in multiple object tracking is associating the data for linking object detections to target. In some studies, this problem is solved using a probabilistic approach [7, 8] or a deterministic approach, for eg: Hungarian algorithm in [9, 10].

In this Mean Shift algorithm was employed to track the vehicles but the short-comings of this algorithm led to the implementation of 8-point connectivity. The execution of this algorithm was done on the Carla simulator, but this algorithm works on any given semantic segmented video as a input.

An effective collision avoidance system can be developed by multiple vehicle tracking which gives information to the ego vehicle with respect to surrounding objects and vehicle. With vehicle tracking safe navigation of autonomous vehicles can be achieved.

2 Related Work

2.1 Multiple Road-Objects Detection and Tracking for Autonomous Driving

Author: Wael A. Farag The American University of the Middle East.

An autonomous driving real-time road object detection and tracking (LR ODT) method is put forth in this paper [15] by the author. The employed technologies is centered on the combination of lidar and radar measurement data [11], where the sensors are put on the same automobile and their data fusion is employed on unscented Kalman filter (UKF). The fusion approach applied by them offer both pose and speed data for obstacles moving in roadways surrounding the ego automobile. The primary contribution of this study is the balanced handling of pose estimate precision and its real-time performance. The proposed method makes use of optimized math and optimization libraries and is implemented in high-performance language C++ for high real-time performance. The simulation experiments were conducted to evaluate how successfully the LR ODT tracks bi-cycles, autos, and people. The UKF fusion's performance is contrasted with that of the extended Kalman filter fusion (EKF), like in approach [12], performance demonstrating its superiority. On all test scenarios and state variable levels, the UKF has beaten the EKF (−24% average RMSE). The tracking performance gain over using a single device is remarkable, as demonstrated by the fusion technique which has −29% RMES with lidar and −38% RMSE with radar.

2.2 Automated Multi-object Tracking for Autonomous Vehicle Control in Dynamically Changing Traffic

Author: Beshah Ayalew, Qian Wang.

In this paper [13] author uses automatic multi-object tracking algorithm. The method can handle missed targets as well as target presence and disappearance, and it was integrated with an MPC-based steering algorithm, as discussed in [14], for a driverless car application. They have demonstrated how an LMIPDA-based tracking management system independently activates/terminates, tracks and provides track continuity under transient missed detections using simulated real-world traffic circumstances. They also demonstrated how the MPC module leverages tracking data from the tracking component to make more informed judgments. Furthermore, a motion prediction approach is required for autonomous navigation in public traffic to follow the target automobiles' future intentions. The IMM subsystem predicts the target cars' N-step forward manoeuvre in this investigation, which restricts the MPC module's optimization issue.

3 Methodology

To perform vehicle tracking, the data set is taken from the Carla simulator. Multiple vehicle tracking and distance calculation is achieved with the fusion of two camera sensors i.e Semantic camera and depth sensor camera.

3.1 Objectives

- Understanding the problem statement with identification of boundaries to provide a technical solution.
- To propose different Algorithm for vehicle tracking with respect to ego vehicle for different scenarios
- Develop a mathematical model for tracking and test it on different data set.
- Verification and validation of the model using suitable simulation tool.

3.2 Specifications

The implementation of the tracking algorithms is carried out in Carla simulator. Some of the libraries that have been installed in python are Open CV, Pygame, NumPy. The commands to run the program are given as follows:

- CarlaUE4.exe -windowed-Carla-server -Carla -word-port = 2003
- Python.exe manual control.py –port = 2003
- Semantic camera has been used for vehicle detection
- Depth camera has been utilized for distance calculation.

3.3 Architecture

Detecting several elements in a picture is a difficult operation; nevertheless, with current approaches, it is possible to do so with precision in real time. The object detection is carried out using feature extraction by using different algorithms. The implementation of tracking was done by the virtue of two algorithms. The drawbacks of mean shift algorithm was solved using 8-point connectivity. Each detected vehicle is given a bounding box with a size of 3 units.

Single Vehicle Tracking. Single vehicle tracking is carried out with the mean shift algorithm. This was introduced in computer vision by Cheng [16]. The author has briefly explained about Mean shift. It's a algorithm that transfers a data point to the average of sample points in its vicinity. Useful for grouping, mode searching, estimating probability density and tracking. The algorithm outline is given in Fig. 1.

Working:

- Begin with datasets that have been allocated to their respective cluster.
- This method will then compute the centroids.
- The position of new centroids will be updated in this stage.
- The procedure will now be iterated and shifted to a greater density zone. Finally, it will be terminated when the centroids reach a point where they can no longer advance any further.

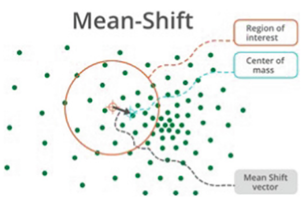


Fig. 1. Algorithm outline

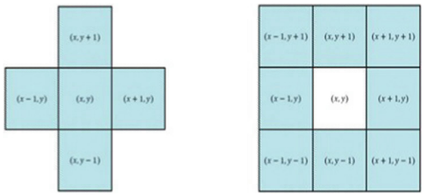


Fig. 2. 8-point and 4-point neighborhood

Multiple Vehicle Tracking. The 8-point connectivity algorithm is used to monitor multiple vehicles. To recognise objects in a digital image, we must find groupings of pixels that are "linked" to each other. In other words, the objects in a particular digital image are the linked components of that pattern. If two pixels, X and Y, have a common edge or a vertex, they are 8-neighbors (or simply neighbours). The Moore neighbourhood of a particular pixel X sis made up of its 8 neighbours. 8 connected pixels are neighbours to every pixel that touches one of the edges of the corner Vertically, horizontally and diagonally. The neighbourhood of both 8-point and 4-point connectivity is shown in Fig. 2. Using 8-point connectivity we are clustering pixel coordinates of different vehicles.

3.4 Block Diagram

Figure 3 briefly depicts the basic steps of the algorithm. The semantic camera first records a video sequence and gives the system semantic image input. Later, it is divided into three RGB channels, where the pixel values of each image are assessed. The image with the highest pixel value is then taken into consideration and given to our tracking algorithm, which is further processed to give us the output by tracking the multiple vehicle with reference to our ego vehicle.

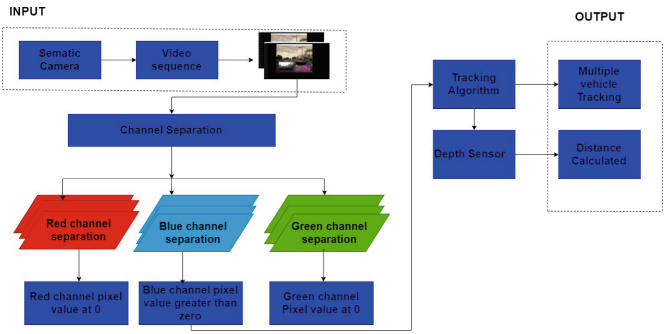


Fig. 3. Functional block diagram

3.5 Mathematical Modelling

Mean Shift: Mean shift is one of the clustering algorithms which assigns the data points to the clusters by iteratively shifting the points to the mean. The algorithm iteratively allocates each data point to the nearest cluster centroid given a collection of data points; the location to the closest cluster centroid is decided by where the majority of the adjacent points are located. Each data point will therefore get closer with each iteration to the location of the majority of points, which is or will be the cluster center. Each point is given a cluster after the algorithm ends. The number of clusters need not be predetermined for mean-shift. The program evaluates the input and determines the number of clusters.

"Mean shift" is a repeated process of "moving to the mean". Every data point in the algorithm is gradually migrating to the "regional mean", and the position of each point's eventual destination reflects the cluster to which it belongs.

Consider a sample data set for clustering as shown in Fig. 4. The mean point in this context is defined as the arithmetic mean of all the features, because it is determined using equal weights for all points. $M = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1+x_2+\cdots+x_n}{n}$

Mean can also be calculated using the weights. The weighted mean is calculated as, $M = \frac{1}{n}\sum_{i=1}^{n} x_i w_i$.

M is the mean, n represents the size and x is the features of the data points, w is the weight for each feature x. To establish the window area size, the bandwidth is set to 2. The middle points for each cluster region is calculated and finally the clustered points are separated with different colors.

The result of mean shift algorithm on the sample data set is shown in Fig. 5.

8-Point Connectivity:
G(a,b) is used to represent an image, while x,y is used to represent each individual pixel. A pixel at (a,b) has four neighbors in the horizontal and vertical directions: (a+1,b), (a−1,b), (a,b+1), and (a,b−1). These are referred to as X's N4 neighbors (X). The four diagonal neighbors of a pixel X at (a,b) are (x+1,y+1), (x+1,y−1), (x−1,y+1), and (x−1,y−1). These are referred to as X's diagonal-neighbors: ND (X). The 8-neighbors of X are the 4-neighbors and the diagonal neighbors: N8 (X).

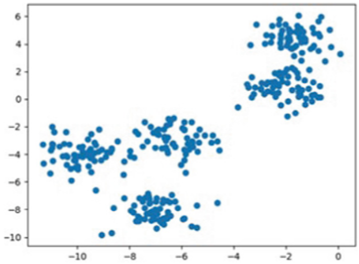


Fig. 4. Sample data set

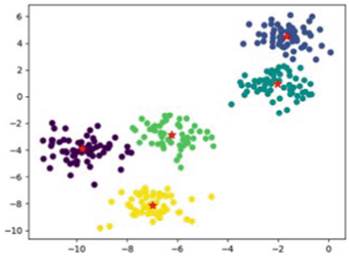


Fig. 5. Algorithm implementation on data set

3.6 Algorithm

Mean Shift Algorithm

	Red channel	
50	250	250
0	215	215
215	215	215
0	215	18

	Green channel	
5	0	0
210	0	0
0	0	0
0	0	0

	Blue channel	
5	210	0
0	0	0
0	0	0
15	0	0

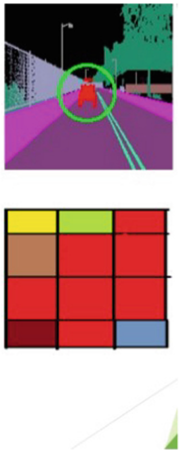


Fig. 6. Understanding the algorithm using dataset

- Looking at the image pixels in Fig. 6, we need to filter with respect to the region of interest and extract the required pixel values.
- Image pixels should fall under condition to be part of the target object. Condition: Red pixel value should be greater than 215 and blue pixel value should be equal to 0 and green pixel value should also be equal to 0.
- After applying the filter, the coordinates will be: [(0,2), (1,1), (1,2), (2,0), (2,1), (2,2) (3,1)]
- Now, by applying mean shift algorithm, the centre of the target can be found.
  Initial centroid: (2,0)
  Window shape: circle
  Window size: 3
  Distance from centroid: $(x_1 - c_x)^2 + (y_1 - c_y)^2(0.5)$
- When the mean shift algorithm is applied, the new centroid will be the mean of the points whose distance from the present centroid will fall under the window size.

8-Point Connectivity

- Fetch all the blue pixels coordinates present in the image (semantic image converts vehicle blue).
- Blue pixel coordinates = [(0,3), (1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,1), (3,5), (3,8), (3,9), (4,7), (4,8), (4,9), (5,7), (5,10)].
- Checking 8-point connectivity for (0,3):
  [(0,3), (1,3), (1,2), (1,4)] : Satisfied
- Checking 8-point connectivity for (1,2): Satisfied
- Checking 8-point connectivity for (1,4): Satisfied
- Calculating centroid of all the vehicles classified by step 2.
- Map the centroid to the depth sensor image.
- Calculate the distance from the ego vehicle
  Distance(meters) = [1000 * (R + G * 256 + B * 256 * 256)/(256 * 256 * 256 − 1)]

4 Implementation

4.1 Platform Details

This project is being carried out on the Carla simulator platform. CARLA stands for CAr Learning to Act which is an open source autonomous driving simulator. It is built from scratch to support for training and prototyping it has a modular API that is used to address issues with autonomous vehicles. CARLA runs over the Unreal engine which provides the feature of urban environment using OpenDrive standard to define the road and other urban settings. It also provides various free real time digital content that can be used like urban layout, pedestrians on the road, buildings and different vehicle models. It also has option to customize different weather like (rainy, sunny) to test in different weather conditions. The highlighted features of CARLA are as follows:

- Traffic Manager: The client side module built on the top of Carla C++ API is responsible for controlling the vehicles in the simulation and populates the road with multiple vehicle.
- Scenario Runner: It is used for developing the traffic scenarios in CARLA. These scenarios are defined using Python interface.
- Responsibility Sensitive Safety (RSS): A mathematical model is used by a C++ library that is connected with CARLA to look into the behaviour of RSS.
- Semantic Segmentation: Semantic segmentation is a technique used by autonomous cars to help them find their way. This technique for understanding the environment divides a picture into sections that belong to the same object by classifying each and every pixel in the image.

4.2 Optimization of Algorithm

In the algorithm, centroid for first iteration was the first car's pixels detected. Since mean shift algorithm is an iterative process, it would create a delay in the convergence of the centroid. So optimization was done by setting the centroid detected in previous frame as the centroid for first iteration, which increased speed of convergence by reducing the number of iterations.

For optimization in multiple vehicle tracking, 4-point connectivity was implemented to facilitate multiple tracking. It would look for neighbouring pixels from a list containing pixel coordinates of all the vehicles. In this pattern, pixels were classified and number of vehicles was detected. 8 point connectivity was substituted in place of 4-point connectivity which decreases the number of iterations and increases the accuracy of multiple vehicle tracking.

Shifting from mean shift algorithm to 8-point connectivity: Mean shift algorithm successfully accomplished the task of single vehicle tracking but lacked in the area of multiple vehicle tracking. 4 point connectivity is efficiently capable of multiple vehicle tracking but the computational speed is slow and the algorithm was not up to the mark. So, using 8 point connectivity instead of 4 point connectivity made the algorithm more robust and accurate.

5 Implementation Results

After implementing single vehicle tracking, the number of vehicles approaching in the field of view was unknown. While trying to track multiple objects, the convergence of centroid was also a major issue. There was a problem of initialisation of the centroid. This issue was solved using 8 point connectivity and multiple vehicle tracking was achieved.

The single vehicle tracking result is shown in Fig. 7. Tracking of surrounding vehicles is done by the 8-point connectivity concept in the Carla simulator. Multiple vehicle tracking can be seen in Fig. 8.

The implementation can also be done using 4-point connectivity but to optimize the algorithm 8-point connectivity has been used. The area that the algorithm fails to track the vehicle is when occlusion occurs. Occlusion can lead the system to keep track of the target being monitored, or it can cause the erroneous object to be tracked after overlapping. When two vehicles are seen very closely to each other the algorithm considers the entire cluster to be one.

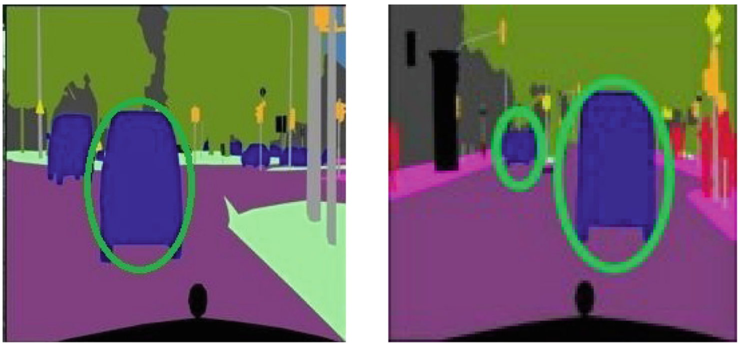


Fig. 7. Single and multiple vehicle tracking

6 Comparison Against Related Work

In this part, we compare our proposed framework to alternative vision-based (with LiDAR or RADAR data) tracking techniques. Generally object tracking involves the use of sensors like Radar and Lidar.

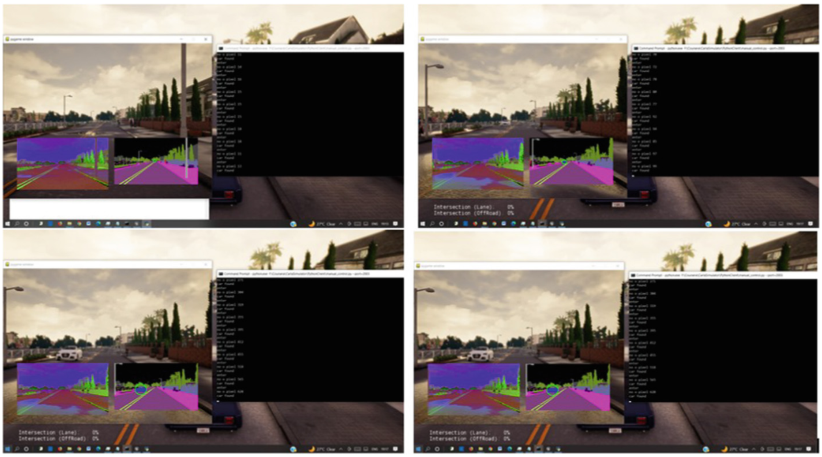


Fig. 8. Implementation results under different scenarios

6.1 Comparison Against Tracking Methods Using Sensors

Accurate image recognition can be obtained for cameras in autonomous driving applications. The accuracy of the 8 point algorithm is higher as compared to the 4 point connectivity. Despite the fact that LiDAR has great measurement accuracy, comprehensive information gathering, and strong real-time performance, it is still unable to fulfil all tasks such as target identification and recognition and environmental modelling. In comparison, given current technical circumstances, LiDAR can better handle the problem of distance measurement and increase the safety of autonomous cars on the road. This is not possible with computer vision.

6.2 Comparison of Image Processing Algorithms Against Machine Learning

Classical Image processing algorithms are more mathematically precise but machine learning algorithms are more data-intensive. Learning models are not cost-effective to devote a large amount of computer resources to learning a very complicated model of a fast changing network environment. ML is likewise more focused on prediction, whereas Statistical Modeling is more focused on interpretation.

7 Conclusion

Tracking of vehicles for driver less cars is critical for safety and accident prevention. The study describes how images may be a valuable resource for detecting and classifying items in the vicinity of a vehicle. Main target of the project is to determine and track the exact trajectory path and direction of the obstacle/car which is in the FOV (field of view). The user will get the exact number of cars approaching towards the ego vehicle and will also be continuously updated about the path of the obstacle. Accurate distance of the obstacle will also be notified to the user.

8 Future Scope

The only scope exclusive in the project is the problem of occlusion. The main future scope of the project will be to solve the problem of occlusion which will make the implementation more precise and efficient.

References

1. Cho, H., Seo, Y.-W., Kumar, B.V., Rajkumar, R.R.: A multisensor fusion system for moving object detection and tracking in urban driving environments. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 1836–1843. IEEE (2014)
2. Allodi, M., Broggi, A., Giaquinto, D., Patander, M., Prioletti, A.: Machine learning in tracking associations with stereo vision and lidar observations for an autonomous vehicle. In: Intelligent Vehicles Symposium (IV), pp. 648–653. IEEE (2016)
3. Pfeiffer, D., Franke, U.: Efficient representation of traffic scenes by means of dynamic stixels. In: Intelligent Vehicles Symposium (IV), pp. 217–224. IEEE (2010)
4. Osep, A., Hermans, A., Engelmann, F., Klostermann, D., Mathias, M., Leibe, B.: Multi-scale object candidates for generic object tracking in street scenes. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 3180–3187. IEEE (2016)
5. Satzoda, R.K., Lee, S., Lu, F., Trivedi, M.M.: Vision based front rear surround understanding using embedded processors. IEEE Trans. Intell. Veh. **1**(4), 335–345 (2017)
6. Sivaraman, S., Trivedi, M.M.: Dynamic probabilistic drivability maps for lane change and merge driver assistance. IEEE Trans. Intell. Transp. Syst. **15**(5), 2063–2073 (2014)
7. Khan, Z., Balch, T., Dellaert, F.: MCMC-based particle filtering for tracking a variable number of interacting targets. IEEE Trans. Pattern Anal. Mach. Intell. **27**(11), 1805–1819 (2005)
8. Oh, S., Russell, S., Sastry, S.: Markov chain Monte Carlo data association for multi-target tracking. IEEE Trans. Autom. Control **54**(3), 481–497 (2009)
9. Bae, S.-H., Yoon, K.-J.: Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1218–1225 (2014)
10. Kim, S., Kwak, S., Feyereisl, J., Han, B.: Online multi-target tracking by large margin structured learning. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7726, pp. 98–111. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37431-9_8
11. Chavez-Garcia, R.O., Aycard, O.: Multiple sensor fusion and classification for moving object detection and tracking. IEEE Trans. Intell. Transp. Syst. **17**(2), 252–534 (2015)

12. Nanda, J.S., Ghosh, S.: Performance comparison of EKF and UKF for estimation of COE using Kozai mechanism. In: 2021 2nd International Conference on Range Technology (ICORT), pp. 1–6 (2021)
13. Hunde, A., Ayalew, B., Wang, Q.: Automated multi-object tracking for autonomous vehicle control in dynamically changing traffic. In: 2019 American Control Conference (ACC), pp. 515–520 (2019)
14. Seo, J.-H., Oh, K.-S., Noh, H.-J.: Model predictive control–based steering control algorithm for steering efficiency of a human driver in all-terrain cranes. Adv. Mech. Eng. **11**(6), 168781401985978 (2019)
15. Farag, W.: Multiple road-objects detection and tracking for autonomous driving. J. Eng. Res. **10** (2021). https://doi.org/10.36909/jer.10993
16. Cheng, Y.: Mean shift, mode seeking, and clustering. IEEE Trans. Pattern Anal. Mach. Intell. **17**(8), 790–799 (1995)

Sexism Classification in Social Media Using Machine Learning Algorithms

B. Karthikeyan, Srivatsa Sundarraj, Charuni Sampathkumar, K. Mouthami(✉), and N. Yuvaraj

KPR Institute of Engineering and Technology, Coimbatore, India
mouthami.k@kpriet.ac.in

Abstract. Currently, social media is a place where everyone shares their perspective. Sexism mainly affects women which may eventually lead to physical and physiological abuse. It is similar to gender discrimination. For many years, sexism has prevailed in our society and affects social media too. In the proposed paper, the analysis of sexism is done using NLP and Machine Learning algorithms. In the world, male domination has taken a firm root, and women are trying hard to break the stereotypes that women cannot work or are only for giving birth and caring for children. The dataset used is EXIST-2022 and worked only in the English language. The classification uses machine learning algorithms such as Support Vector Machine, Random Forest, Naive Bayes and Logistic Regression. The proposed paper discusses the performance of multiclass and binary classification methods to work with various kinds of sexism on social platforms. The best accuracy obtained for multiclass classification and binary classification.

Keywords: Random Forest (RF) · Natural Language Processing (NLP) · Naive Bayes (NB) · Support Vector Machine (SVM) · Logistic Regression (LR)

1 Introduction

Social networks have seen a tremendous increase in users over the past few years. It has become a part of our world, and it reveals how our society is. With the help of social networks, interaction between people has become fast and straightforward. It has enabled socializing between different cultures, countries, and ethnicities between people. However, social networks have many adverse effects like depression, cyberbullying, addiction, and unhealthy sleep patterns. Our life is based on social media and it is the biggest element present in our life. Social media has vast consequences on mental health. Social association with people can minimize strain, improve self-assurance, provide support and happiness, prevent isolation, and even increase life span. Social media can never be a replacement for face-to-face human interaction and can be a severe threat to your intellectual and sentimental health. Spending much time on social networks may cause a feeling of sadness, dissatisfaction, exhaustion, and frustration. According to the Pew research center, 40% of adults have been a victim of cyberbullying and harassment [1]. Sexism is a significant abuse faced by both genders, mainly females. One should

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 14–23, 2023.
https://doi.org/10.1007/978-3-031-27524-1_2

not spread hatred while expressing their thoughts and communicating on social media. It may affect people in many ways. Sexism reduces the freedom of females to a great extent. One tweet can change the perspective of a person's life and impact their entire life. Males and females should be considered equal, not above or below each other. Everyone should work together to eliminate sexism and harassment in all places. Verbal violence can be reduced if everyone communicates after thinking twice.

The data was gathered from the sexism Identification in Social networks (EXIST2022) dataset, which has annotated messages from social media. The model was classified based on types of sexism like ideological inequality, non-sexism, objectification, sexual violence, stereotyping dominance, misogyny nonsexual violence. The data is classified based on binary sexism classification as sexist and non-sexist messages. Natural Language Processing (NLP) is used to preprocess the dataset. The machine learning algorithms applied are Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) discuss which works best for the classification process. The proposed paper will discuss both binary and multiclass classification. The remainder of the paper is structured as follows: Related works are discussed in Sect. 2. The methodology is provided in Sect. 3. The results and analyses were described in Sect. 4. Section 5 provides the conclusions and suggested next research. Section 6 presents the references.

2 Related Works

The dataset confirms that sexism classification in social media is thought-provoking, with a large room for improvement [2]. The binary classification was done with Spanish twitter data and it segregates the tweets as sexist, on-sexist, and doubtful. This paper takes the Spanish data, converts it into English, and does pre-process based on different types of sexism that target women. The findings demonstrate that sexism regularly appears in social networks in various forms, that a combination of manners is included, and those deep learning algorithms are used to detect them. This paper discusses how well automatic classification techniques function when dealing with various types of sexism to other subdomains, like misogyny. It uses the Bilingual Encoder Representations from Transformers (BERT) algorithm and produces 74% accuracy. Many algorithms like Bidirectional Long Short-Term Memory (BI-LSTM), BERT, and Multilingual BERT (mBERT) were used [3].

EXIST2022 dataset does binary and multiclass classification. Two multiple language transformers and two alternative methods were used to resolve the two associated tasks. Both supervised fine-tuning of sexist content and unsupervised pre-training with additional data provide enriched data. The XLM-RoBERTa model applies to both jobs. The performance of the three models is best when the two techniques are combined. It is getting an F1 score of 74% and 46%. mBERT, T5, and XLM-R were used. Additional datasets were used to pre-train the dataset [4]. Deep decision tree classification improves the accuracy of cyberbullying [5]. Hate speech detection explains how hate speech is prevalent in social media and how to use NLP to detect hate speech. It addresses the challenges in hate speech detection, the models that can be used, and how to improve the performance for observing hate speech. Its abusive things that can be said about a particular individual or a group of people. In sentiment analysis, negative labels can be considered hate speech. So, hate speech detection is about differentiating a hateful speech utterance from a harmless statement. Sentiment analysis is very similar to hate speech itself [6]. Machine learning and NLP were utilized for hate speech classification. It uses emotions to detect speech. It has 80% accuracy while using SVM. It used a lexicon-based analysis method, and it helped to improve accuracy [7].

Gender-based classification of messages in Reddit and analyzing the subreddits [8]. Dividing the comments on YouTube, Reddit, Wikipedia, and Twitter into 20% of hateful comments and 80% of non-hateful comments. Then use several classification algorithms like SVM, LR, NB, XGBoost and Neural Networks and feature representations like Bag-of-Words, Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, BERT, and their combination and XGBoost get an F1 score of 0.92 as the maximum [9]. Improved RBF kernel of SVM with 98.8% accuracy for sentiment analysis [10]. A novel deep learning system for multiclass classification of tweets with a 0.59 accuracy rate [11].

3 Methodology

3.1 Dataset Identification

The dataset used was the EXIST2022 benchmark dataset and it has 11,345 annotated social network sentences. It was separated into testing and training datasets. It had multilingual data and only English data was utilized. The training set consists of 10210 values. The testing data consists of 1135 values. Each data instance is given a binary label to show whether the message is non-sexist or sexist. Also, a multi-class arrangement is given: ideological inequality, non-sexist, objectification, sexual violence, stereotyping dominance, and misogyny nonsexual violence (Tables 1 and 2).

Table 1. Binary classes and their examples

Class	Example
Sexist	I HATE FB someone said “where the real women at” and in the comments someone else said “church” and a dude responded with “every girl i met at church was a whore”
Non-sexist	@BankofIndia_IN Devil Branch, New Delhi should be avoided by all and deserves being shut because the employees therein are least bothered with the consumer’s woes! 20 days yet no redressal of woes! #disappointed

3.2 Preprocessing

Preprocessing is converting data into an understandable format. Preprocessing will improve completeness, consistency, and interpretability. Preprocessing is an integral

Table 2. Multiclass classes and their examples

Class	Example
Non-sexist	@Banyosssss My philosophy: He has a funny mustache what else does a man need
Objectification	@CurvyBandida @Xalynne_B Wow, your skirt is very short. What is it's length? 5 inch or more?
Stereotyping dominance	I genuinely want to be wealthy, but not wealthy like a trophy wife, wealthy on my own
Misogyny nonsexual violence	@Sshania03 Im a Sagittarius so I hate Sagittarius women

part of NLP as a dataset may have noise, affecting the model's accuracy. So, preprocessing will be an essential step in text cleaning. Regex was used to remove the texts that were not necessary for the model's training. Regex means regular expressions, a sequence that helps to find patterns in a sentence. The regex removes the hashtags, mentions, links, non-ASCII characters, punctuations, and white spaces. Then tokenizer is used to convert the sentences into words. The stop words are removed from the tokens.

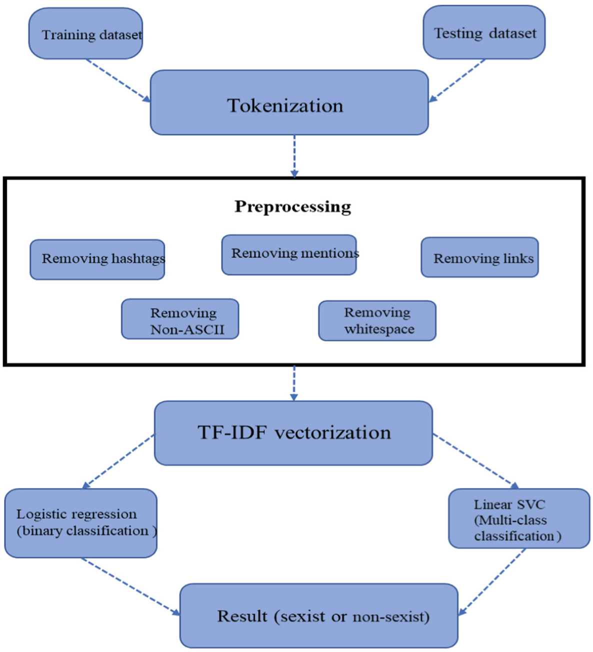


Fig. 1. Proposed methodology for the binary and multiclass classification

Figure 1 represents the methodology that was followed for binary and multi-class classification. It follows tokenization, preprocessing, vectorization, and model implementation.

3.3 Models

The Machine Learning models used for classification are Support Vector Machine, Logistic Regression, Random Forest and Naive Bayes. The models section will discuss the methodology of each algorithm in detail.

Logistic Regression. The algorithm is a variety of regression methods where a decision threshold is used and based on the threshold, the model classifies the output as 0 or 1 (binary classification). Setting a proper decision threshold is essential. Precision and recall affect the threshold value. It can be applied to binary and multiclass types. It models the data using the sigmoid function mentioned in Eq. (1) [12]. The value of the sigmoid function lies between zero and one.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

It represents a mathematical function having a characteristic "S"-shaped curve or sigmoid curve. Sigmoid curves are commonly used in statistics as cumulative distribution functions and as integrals of the logistic density, the normal density, and Student's t probability density functions. The logistic sigmoid function is invertible, and its inverse is the logit function.

Linear SVC. SVM is a supervised learning algorithm for regression, classification, and outlier detection. Here the SVM Classification (SVC) method was used [13]. SVM chooses the extreme points that help in producing the hyperplane. It selects a decision boundary that maximizes the distance from all the classes' nearest data points. The boundary in SVM is called the maximum margin classifier or the maximum margin hyperplane. Linear SVC attempts to find a hyperplane to maximize the distance between classified samples shown in Fig. 2.

Multinominal NB. The algorithm is used for text classification of multi-scale data. It assists in developing a quick machine-learning model that can create rapid predictions. It uses a probability-based classifier to classify the data [14]. The NB algorithm works depending on the Bayes rule. It can be used for both multi-class and binary classifications. Multinomial NB is a popular machine learning algorithm implemented to analyze categorical text data. The NB is a group of multiple ML methods that will share a common principle, and one feature is independent of every other one present. The existence or absence of one character has no bearing on the other feature's presence or absence.

Random Forest Classifier. The algorithm contains a vast number of single decision trees operating together as a bagging ensemble model. Ensemble learning models are

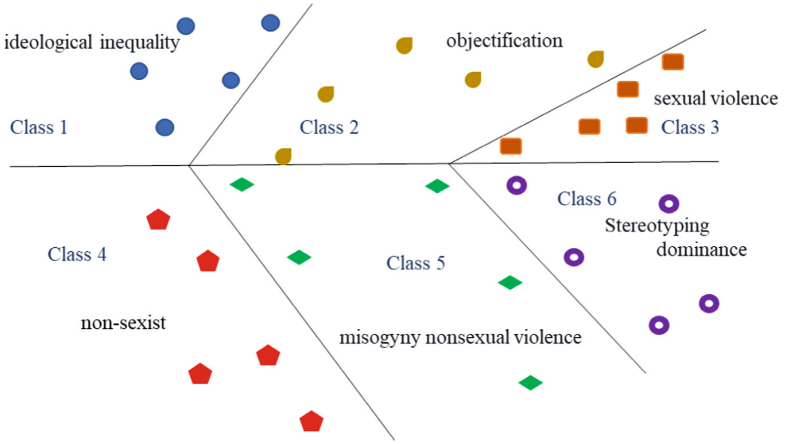


Fig. 2. Multiclass support vector classification model

made up of a set of classifiers. The idea behind decision trees is that you use the features to create yes or no questions and split the dataset continuously until all data points are isolated belonging to each class based on decision gain. The RF algorithm is an extension of bagging-based ensemble learning [15]. Here, the model majorly used is trees. The RF model creates decision trees from an arbitrarily selected subset of the dataset. Each tree in the forest algorithm gives out a prediction class. The class which has the majority vote becomes the prediction result [16].

4 Results

In this section, the results acquired by multiple models are studied. The state of the evaluation methods is explained. For calculating the evaluation metrics, the data is split up into 90% training and 10% validation set, giving better and improved results than the other split-ups. For binary classification, a macro average F1 score of 75% is the best. The binary categorization has a 74% average mean accuracy. For binary classification, LR performs well. Table 3 and Fig. 3 will show us the performance metrics for the binary classification.

The confusion matrix shows how the LR model for binary classification has predicted the true positives and the false positives **are shown** in Table 4.

Table 3. F1 score, recall, support, and precision for binary classification

	Precision	Recall	F1 score	Support
Linear SVC	0.71	0.71	0.71	1129
RF classifier	0.72	0.68	0.67	1129
LR	0.74	0.74	0.74	1129
Multinomial NB	0.72	0.72	0.72	1129

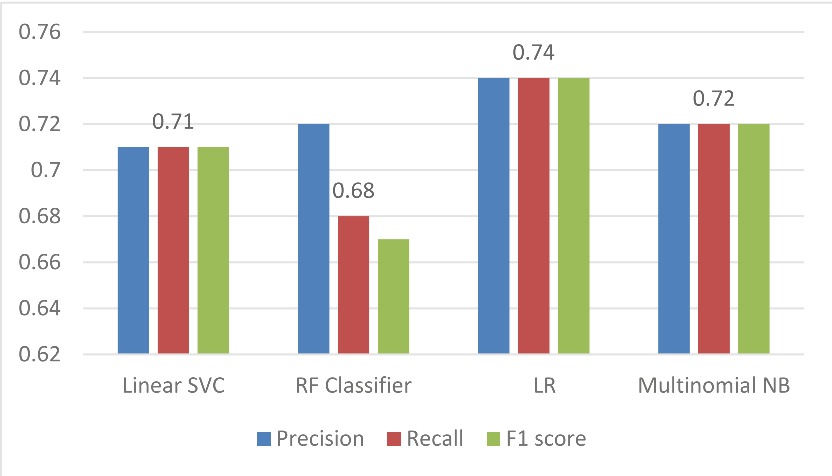


Fig. 3. Performance evaluation of binary classification

Table 4. Confusion matrix for binary classification

	True	False
True	424	142
False	150	413

The best macro average F1 score for multiclass classification is 50%. The average mean is 62% for the multiclass classification. The binary classification works better than the multiclass classification. Linear SVC works well for multiclass classification. Table 5 and Fig. 4 will show the performance metrics for the multiclass classification.

The multiclass classification model works by checking each class and its accuracy and precision for the Linear SVC for each class are present in Table 6 and Fig. 5.

Table 5. F1 score, recall, support, and precision for multiclass classification

	Precision	Recall	F1 score	Support
Linear SVC	0.56	0.46	0.50	1129
RF classifier	0.08	0.17	0.11	1129
LR	0.67	0.40	0.46	1129
Multinomial NB	0.69	0.24	0.24	1129

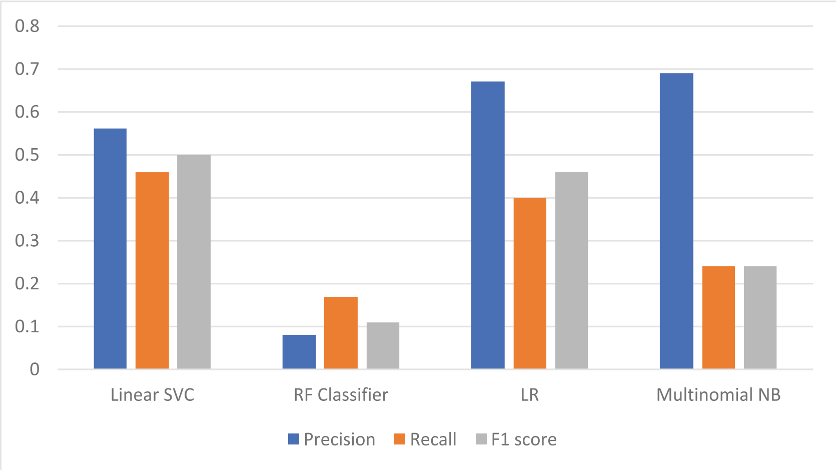


Fig. 4. Performance evaluation of multiclass classification

Table 6. The precision and recall multiclass classification

	Precision	Recall
Ideological-inequality	0.74	0.42
Non-sexist	0.60	0.27
Objectification	0.62	0.95
Sexual-violence	0.73	0.18
Stereotyping-dominance	0.69	0.25
Misogyny-non-sexual-violence	0.66	0.34

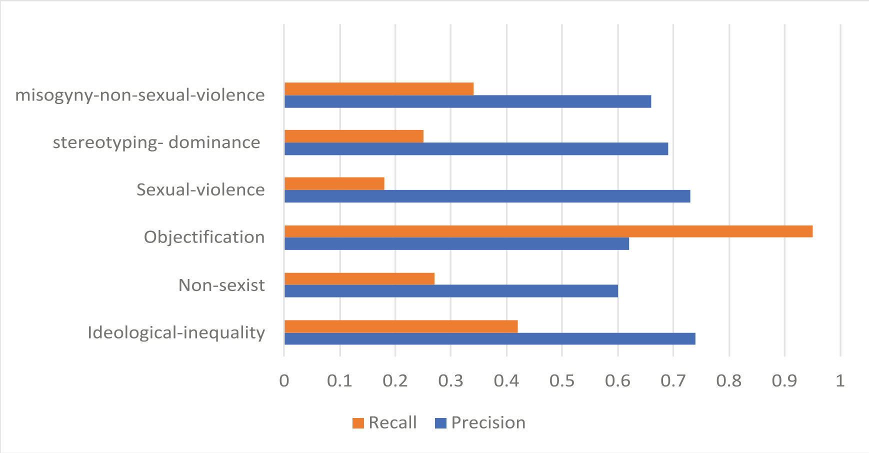


Fig. 5. Performance evaluation of binary classification

5 Discussions and Conclusions

In this paper, the classification is developed for the tweets based on the sexism that prevails in social networks. Online hate has increased significantly over the few years, and it can be classified using the model that has been developed. Social media is a place where there is freedom of speech but one should use it wisely and should not spread hatred.

Table 7. Mean accuracy for both binary and multiclass classification

Binary classification	Mean accuracy
Linear SVC	0.697288
LR	0.715359
Multinomial NB	0.686036
RF classifier	0.615164
Multiclass classification	Mean accuracy
Linear SVC	0.603116
LR	0.595675
Multinomial NB	0.542788
RF classifier	0.504961

Models like RF, NB, SVM, and LR were applied to understand how it works on binary and multiclass classification. Despite the clear benefits and beneficial outcomes of this worldwide communication, the accessibility, invisibility, and anonymity of the internet have made it easier for irrelevant, racist, and one-sided speech is unchecked and undisciplined. The primary goal is to find and understand how sexist frame of mind and conduct are expressed in social network conversations. The EXIST2020 dataset was used with only the English language. Then worked on preprocessing it by removing the significant noise in the data. The models were applied, and the outcome shows that the binary classification performs well than the multiclass classification. In the future, the upgradation of accuracy and implementation for other languages can be done and combined to form a multilingual classification. Additional data can be added to the dataset as more data improves the model's performance (Table 7).

References

1. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/
2. Rodríguez-Sánchez, F., et al.: Overview of exist 2021: sexism identification in social networks. Procesamiento del Lenguaje Natural **67**, 195–207 (2021)
3. Rodríguez-Sánchez, F., Carrillo-de-Albornoz, J., Plaza, L.: Automatic classification of sexism in social networks: an empirical study on Twitter data. IEEE Access **8**, 219563–219576 (2021)
4. Mina, S., et al.: Automatic sexism detection with multilingual transformer models. arXiv preprint arXiv:2106.04908 (2021)
5. Yuvaraj, N., et al.: Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification. Comput. Electr. Eng. **92**, 107186 (2021)
6. Schmidt, A., Wiegand, M.: A survey on hate speech detection using natural language processing. In: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, pp. 1–10 (2017)
7. Raffel, C., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. **140**, 5485–5551 (2020)
8. Martins, R., Gomes, M., Almeida, J.J., Novais, P., Henriques, P.: Hate speech classification in social media using emotional analysis. In: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS), pp. 61–66 (2018)
9. Thelwall, M., Stuart, E.: She's Reddit: a source of statistically significant gendered interest information. Inf. Process. Manage. **56**(4), 1543–1558 (2019)
10. Salminen, J., Hopf, M., Chowdhury, S.A., Jung, S., Almerekhi, H., Jansen, B.J.: Developing an online hate classifier for multiple social media platforms. Human-Centric Comput. Inf. Sci. **10**(1), 1–34 (2020)
11. Gopi, A.P., Jyothi, R., Narayana, V.L., Sandeep, K.S.: Classification of tweets data based on polarity using improved RBF kernel of SVM. Int. J. Inf. Technol. **11**, 1–16 (2020)
12. Jabreel, M., Moreno, A.: A deep learning-based approach for multi-label emotion classification in tweets. Appl. Sci. **9**(6), 1123 (2019)
13. Zou, X., Hu, Y., Tian, Z., Shen, K.: Logistic regression model optimization and case analysis. In: 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), pp. 135–139 (2019)
14. Yin, W., Zubiaga, A.: Towards generalisable hate speech detection: a review on obstacles and solution, PeerJ Comput. Sci. **7**, e598 (2021)
15. Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., Choi, G.S.: A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. Plos One **16**(2), e0245909 (2021)
16. Liakhovets, D., et al.: Transfer learning for automatic sexism detection with multilingual transformer models (2022)

Adaptable Fog Computing Framework for Healthcare 4.0

Karthik Kovuri[1], Katha Chandrashekhar[2], A. V. Sriharsha[3(✉)], Byraboina Siddardha[3], and A. Hitesh Reddy[3]

[1] CSE, B V Raju Institute of Technology, Narsapur, Medak, Telangana, India
karthik.k@bvrit.ac.in
[2] CSE, Indur Institute of Engineering and Technology, Siddipet, Telangana, India
[3] CSE, Sree Vidyanikethan Engineering College, Tirupati, A.P., India
avsreeharsha@gmail.com

Abstract. Healthcare 4.0 is an extension of Industry 4.0. a great leverage has been made to the stakeholders of medical and healthcare domain. Technical innovations and improvisations are exponentially increasing with true significance connected to many industries. Automation and intelligent solutions are the outcome in the trends of industry. All the modifications lead antecedents to the evolving industrial revolution as Industry 4.0. Fog Computing as indispensable ally for the computing framework has predominantly influenced in expansion and development of the Healthcare 4.0. The power of revolution paves healthy waves to the community with Healthcare 4.0.

Keywords: Fog computing · Medical and healthcare systems · Integration · Adoptability · Healthcare 4.0

1 Introduction

Technologies evolve and influence the human life style. The innovations in healthcare industry play predominant role influencing people with economy of infrastructures, accuracy of test results, saving time and centralization or integration of facilities. These luring characteristics have encouraged the scenario of health care industry 4.0, with doors open ajar for the terms intelligence and smart. Smart infrastructures are empowered with Artificial Intelligence, Machine Learning, Deep Learning, Big Data Analytics, Data Science and Internet of Things. Not only in detecting and diagnosing the new ailments that affect the human life style but also in predicting the future health hazards. In the present scenarios of busy world, people are webbed with day-to-day activities and pay very less attention on their regular health. Although medical and health care industry contributes much for propagation and publicity of advantages of modern smart-infrastructures, people neglect to visit and undergo regular medical checkups and yield declination of their health. Governments of various countries offer schemes to enroll and obtain the facilities subsidizing the economy of common man, the attempts go stake. Education with regard to the smart-infrastructures and their applications is also the need of the hour to improve the healthy life style of the people.

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 24–32, 2023.
https://doi.org/10.1007/978-3-031-27524-1_3

The substantial and intelligible factor for medical and health care to upkeep into 4.0, is IoT, smart-infrastructures. Smart-infrastructures are integrate with edge, fog, cloud and a backbone of IoT, offering high scalability, connectivity and availability prudence in performance with high-end big data analytics. Healthcare conditions and medical emergencies need time-critical attention in the territories with communication failures, catastrophe or war calamities. Seamless integration of cloud, fog and edge computing are the right contribution to the stakeholders in timely rendering of healthcare services.

A plethora of technologies influence the world of medical and health care to improve the lifestyle of human beings are very well publicized and outsourced through web and internet, costing less in price and saving lot of time. From the innovative mechanical automation to the trending industry as 1.0, the escalated deployment of intelligent devices is made to 4.0, with many sensible transformations that develop the medical and healthcare along the way to ease down the organizational stress and enhance the capabilities in the establishment. Many challenges lay the fortification of the innovative Healthcare 4.0, particularly in managing the voluminous data vaulted for analyses.

2 Related Work

Smart-infrastructure incorporates smart devices flaming the revolution into the sectors of medical and healthcare. IoT [1] lays the rudimentary roadmap for the development of the smart-infrastructures, collecting data from various ends of the networks. The gamut of the medical and health care systems lies in collecting the data. The collected data plays a significant role in diagnosis, examinations and follow up of treatment. The collected data is more particularly used for prediction of early indications of a disease and provide adequate guidance to overcome the difficult situations. Data is also important for the medical and healthcare campaigns to provide the services at the remote areas through sophisticated technologies like telemedical services.

Therefore, the technology underlying the healthcare 4.0 is extensively the cloud computing technology to store large volumes of data. The speed of transactions and transformation of data is taken care by edge computing approaches where different operational modules are placed at the edge of the network rather at a central place in the cloud. The approach has been proved to solve the issue of high latencies, except large space is required to store the data [2, 3]. Undoubtedly, the size of data requires enormous space of storage on edge devices too, to make the overall technology perform exorbitantly. Representing the complete data for computation on the edge devices is preferred to access the data by the end user. Fog computing is at the devices level, where smart-devices hang on at the tip of the network, connected to the cloud computing as extension of services. Hence, the architecture consists of three major parts cloud layer, edge and fog as represented in Fig. 1.

2.1 Healthcare 4.0

As inspired by Industry 4.0, the medical and healthcare has its successive development into healthcare version of 4.0. The focus congregates on the patient-oriented system with all set augmentation, personalization and virtualization in the medical and healthcare

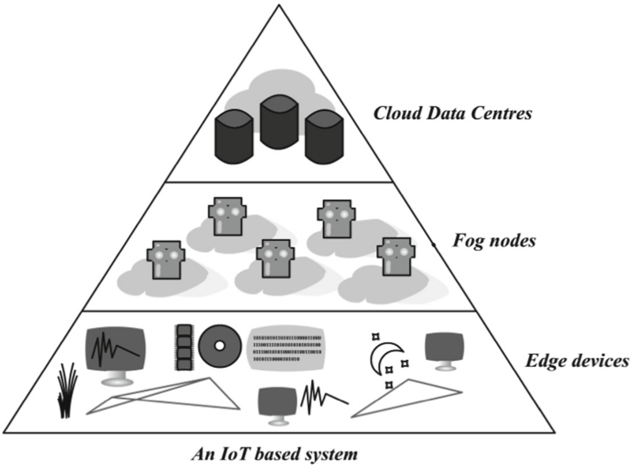


Fig. 1. The consolidated composition of Cloud, Fog and Edge

domain. Maintenance of the Electronic Health Record [2, 3] in the repositories and allowing secured access to doctors without hindrance to location, time and space. The data is shared across the peers and doctors for efficient diagnosis and frame the plan for best treatments. Patients and Healthcare organizations are firmly correlated through technology to given impression no patient is left untreated for any cause of time, space or unavailability of medical resources [4, 5].

2.2 Integration of Fog Computing

Integration of Fog Computing into the Healthcare 4.0 is broadly analyzed and classified into two major areas viz., Patients Health data, Integration of Fog Computing.

2.2.1 Patients Health Data

Majority of data scientists consider this is an important component for the Healthcare 4.0 environment. This aspect can be well mitigated with Data collection and Data analysis [3].

Data Collection: WLAN are widely used in the health surveillance systems with both wearable and non-wearable equipment, which is a very important transformative phase in the healthcare. Low-energy and energy efficient, mobile devices connecting in mobile networks have been advanced exponentially, nonetheless many issues still have to be answered with regard to specific applications and installations. The monitoring systems are the elementary components that first collect the information from the human body as a part of continuous medical surveillance.

Data Analysis: Long term monitoring devices and collected the data needs network space, energy requirements and capacities of the devices. Much of the overhead in the wireless communication can be overcome by means of devices which are utilized for sample collection, compression detection and anomaly-elimination. Bio-monitoring is mandatory for the system using several types of sensor devices fetching sensor information to portals of fog nodes with a specialized wireless communication network.

2.2.2 Fog Computing in Healthcare 4.0

Fog Computing is conceptual when it describes operations connected in Edge Computing, communicating with the processors, programs, storage devices and network resources and further to the data centers [2, 3, 6]. Therefore, the Fog Computing applying more flexibility in conducting the operations at the elementary level of the whole network.

Various approaches of deep learning for image classification and detection in healthcare has been presented in [7–10]. Detection of COVID-19 using Deep Learning methods has been discussed in [11], EEG-Based Brain-Electric Activity Detection During Meditation Using Spectral Estimation Techniques were found in [12], methods for Quality Improvement of Retinal Optical Coherence Tomography shown in [13], Detection of Pneumonia Using Deep Transfer Learning architectures in [14] and Analysis of COVID-19-Impacted Zone Using Machine Learning Algorithms was presented in [15]. Support Vector Machine Classification of Remote Sensing Images with the Wavelet-based Statistical Features was presented in [16]. Convolutional Neural Network is used for detection and classification [17, 18], Detection of Intracranial Haemorrhage [19], COVID-19 Isolation Monitoring [20], Detection of Diabetic Retinopathy in Retinal Images [21], Predictive Analytics for Control of Industrial Automation [22], Robotic Applications [23], Dengue Outbreak Prediction [24] and Protection and Security in Fog Computing [25].

3 Comparisons

The estimation of the systems performance is versatile in Fog Computing, through low energy networks, maintaining low latency through making the intermediate devices available to handle the resilience. Very important aspects are considered for appreciating the competence of the Fog Computing in Healthcare 4.0.

- Latency: Operations on resources stored locally rather directly connecting to the services, in order to eliminate the lower latency issues.
- Capacity: Much sufficient storage space is available in Cloud Computing, Fog Computing uses shared storages of the Cloud Computing and a wide range of data is accessed through network into the Nodes of Fog Computing.
- Bandwidth: Fog Computing works with a reasonable bandwidth that it will not hinder the overall bandwidth of the whole network.
- Responsiveness: As the response time of the Cloud Computing is lower than the response required for the application, the Fog Computing Nodes will render the outputs if they are cached by applications provided in the edge platform.
- Security: A better security is provided in the Fog Computing level, where data is never directly retrieved, accessed for the end-users, therefore providing reliable authentication at the Fog layer.

- Speed: The speed of the transmission of data is high in Fog Computing as compared to any area of the whole network, as minimal data transactions and transformation are encouraged at the elementary level. And comparisons shown in Table 1.

Table 1. Comparison of Fog Computing with other direct applications

Features	Cloud	Edge	Fog
Latency	High	Low	Medium
Capacity	High	Low	Medium
Bandwidth	High	Low	Medium
Responsiveness	Low	High	High
Security	High	Very high	High
Speed	Low	High	High
Data sharing	High	Low	Medium

3.1 Generations of Healthcare

The term healthcare upholds the improvisations of health in terms of diagnostics, methods of treatments mitigation with diseases, preventions and much more importantly the alertness. The upgradation of technologies and assertion of crucial methods in medical and healthcare is composed into versions from time to time, while the revolution in development of industry.

Healthcare 1.0 is the first of its kind; emerged as a rudimentary industry setup where the system is built by the people. Patient and Doctor are the key components of the systems. A direct one-to-one approach is the community-based approach. Technology is incorporated at the level of diagnosis and treatment only. There were individual repositories, even non-digital forms representing the history of the patient. The efficacy of the treatment and diagnosis depends on the patients' ability to narrate the situations and understandability based on the knowledge and experience of the doctor. Just a basic thumb rule of art-and-science of diagnosis and treatment is followed by the doctors.

Healthcare 2.0 is the technology initiative, where technology is a facilitator and collaborator for the medical and health diagnosis and treatments. Application software of social software kind was popularly used as a bridge between doctors and patients in maintaining the electronic health records (EHR). The applications of information technology is maintain the relationship between the doctor and patients through media and ensure the quality-care, competitive-affordability, coherence-in-diagnosis, cost-to-cost services. In 2.0, social networking, substantial understanding of patients about their ailments through documentation, opinion openness and mediating-means. Patients are educated more about their wellness and taking their own healthcare decisions.

Healthcare 3.0 is a value based model, a patient-centric-system, where all the best of the previous versions are consolidated. More towards automation and less administrative hurdles of the institutions, utilizing the opportunities to explore capabilities of expert doctors and engage more in patient-doctor interactions. Healthcare systems are empowered by technology with automated diagnosis rather enslaving on medical evidences. Consequently, medicine is not restricted to the knowledge, but the individual experiences. Therefore, striving towards ascertaining the quality medical services to the patients through sufficient information communication technologies, databases and overcoming the latency in communications.

Healthcare 4.0 is inspired totally by Industry 4.0. A complete conglomeration of the patient-centric components enriched with virtualization and personalization. Electronic Health Records are stored centrally and shared to all the stake holders in the medical and healthcare institutions. Sharing data improves the data accessibility among the doctors irrespective of geographic locations. Patients' data can be reached to expert peers in the doctors' community for ideal suggestions against specific and peculiarities of diagnosis methods and treatments and planning of mitigations with ailments. The strong correlation of patients and doctors through technology in Healthcare 4.0 enables hassle-free outcomes and decisions of treatments and plans. However, challenges related to security, unified-collection-posts, owner-ship-contingencies and communication protocols persist inherently, which shall be solved with research.

3.2 A Comprehensive Study of Healthcare

Healthcare industry is supported by many innovative technologies, methods and frameworks since decades. Electronic Health Record systems are networked to maintain global versions of many important case studies. Healthcare industry is equipped by Artificial Intelligence, Portable Equipment and Real-Time data with improved analytics, shown in Fig. 2. Future scope is very wide for the healthcare industry with many changes being incorporated. Nonetheless, to realize the scope of the healthcare industry, certain important steps are concerned for the development and understanding the potential impacts.

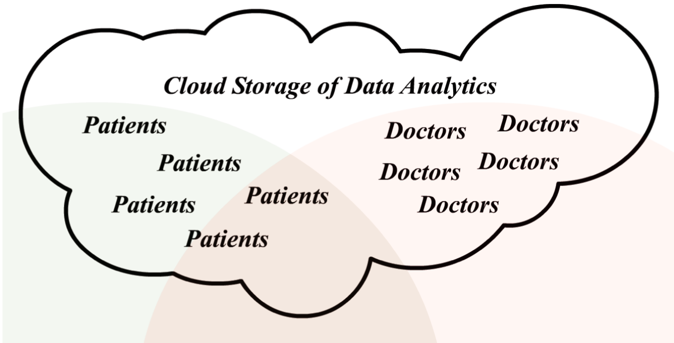


Fig. 2. Healthcare 4.0

3.3 Electronic Health Records

The key component in majority of scientific achievements related to fog computing in healthcare 4.0 is the maintenance of electronic health records and classification of Infrastructure levels in Fig. 3.

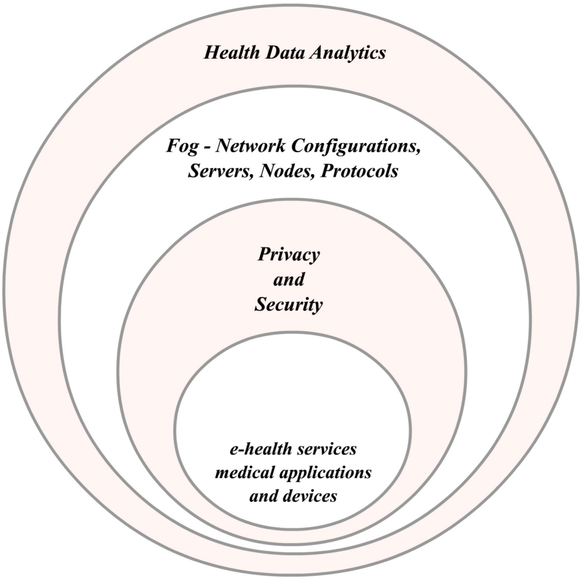


Fig. 3. Classification of Infrastructure Levels in Healthcare – with Fog Computing

Data Collection: The health monitoring and surveillance devices in the networks and wearable forms of medical equipment are the fundamental transformative agents in the health care. Continuous monitoring and van guarding safety in wireless networks have paved opportunities of research exponentially and faired flexible arms in healthcare industry. Low-power devices, energy efficient devices and mobile networks have advanced into sophistication laying at the boundaries of the frameworks collect all important information from the human body and medical equipments.

Data Analysis: The capacities of the devices prove the capacity of collection of data and the accuracy or the width of the data. Differences prevail between the long-term continuous monitoring systems and space and energy requirements of the networks. Samples are collected at the nodes, formatted; transformed, compressed, anomalies are corrected. The wireless communication in the environment is meted with a very less overhead of transportation of the data. Wireless bio-monitoring devices, sensors are defined with specific protocols in the wireless network. The devices, network and the protocols are seamlessly connected to orchestrate a smooth play of collecting, storing and analyzing data.

4 Conclusion

Fog Computing has emerged as a feasible and effective solution to meet various challenges of integrating with Healthcare 4.0. The study of the work, is aimed at drawing the outline and throw light on the implementations and applications of Fog Computing in the medical and healthcare systems. A comparative study of various other technologies have been discussed, which also votes for the Fog Computing as the best resourceful technology to get integrated into the Healthcare 4.0. Compared to previous versions of Healthcare, the version 4.0 shall be the ideal and seamless framework transpired with Fog Computing technology.

References

1. Kumari, A., Tanwar, S., Tyagi, S., Kumar, N., Maasberg, M., Choo, K.K.R.: Multimedia big data computing and Internet of Things applications: a taxonomy and process model. J. Netw. Comput. Appl. **124**, 169–195 (2018)
2. Panigrahi, A., Bhutia, S., Sahu, B., Galety, M.G., Mohanty, S.N.: BPSO-PSO-SVM: an integrated approach for cancer diagnosis. In: Peter, J.D., Fernandes, S.L., Alavi, A.H. (eds.) Disruptive Technologies for Big Data and Cloud Applications. LNEE, vol. 905. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-2177-3_53
3. Sriharsha, A.V., Prema, K.: Differential privacy in big data analytics for haptic applications. Int. J. Compu. Eng. Technol. **8**(3), 11–19 (2017). ISSN: 0976-6375
4. Gupta, R., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M.S., Sadoun, B.: HaBiTs: blockchain-based telesurgery framework for healthcare 4.0. In: 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–5. IEEE, Washington, DC, August 2019
5. Srivastava, A., Singh, S.K., Tanwar, S., Tyagi, S.: Suitability of big data analytics in Indian banking sector to increase revenue and profitability. In: 2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall), pp. 1–6. IEEE, Washington, DC, September 2017
6. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. **2**(1), 3 (2014)
7. Kavitha, T., et al.: Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images. Interdisc. Sci. Comput. Life Sci. **14**, 113–129 (2021). https://doi.org/10.1007/s12539-021-00467-y
8. Madhavi, K.R., Sunitha, G., Avanija, J., Raju, S.V., Abbagallae, S.: Impact analysis of hydration and sleep intervention using regression techniques. Turk. J. Comput. Math. Educ. (TURCOMAT) **12**(2) (2021)
9. Yadlapalli, P., Madhavi, R., Sunitha, G., Jangaraj, A., Kollati, M., Kora, P.: Breast thermograms asymmetry analysis using Gabor filters. E3S Web Conf. **309**, 01109 (2021). https://doi.org/10.1051/e3sconf/202130901109
10. Kora, P., et al.: Automatic segmentation of prostate cancer using cascaded fully convolutional network. E3S Web Conf. **309**, 01068 (2021)
11. Madhavi, K. R., Madhavi, G., Krishnaveni, C. V., Kora, P.: COVID-19 detection using deep learning. In: Abraham, A., Hanne, T., Castillo, O., Gandhi, N., Nogueira Rios, T., Hong, T.-P. (eds.) HIS 2020. AISC, vol. 1375, pp. 263–269. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-73050-5_26

12. Kora, P., Rajani, A., Chinnaiah, M.C., Madhavi, K.R., Swaraja, K., Meenakshi, K.: EEG-based brain-electric activity detection during meditation using spectral estimation techniques. In: Jyothi, S., Mamatha, D.M., Zhang, Y.-D., Raju, K.S. (eds.) Proceedings of the 2nd International Conference on Computational and Bio Engineering. LNNS, vol. 215, pp. 687–693. Springer, Cham (2021). https://doi.org/10.1007/978-981-16-1941-0_68
13. Rajani, A., Kora, P., Madhavi, R., Jangaraj, A.: Quality improvement of retinal optical coherence tomography, pp. 1–5 (2021). https://doi.org/10.1109/INCET51464.2021.9456151
14. Reddy Madhavi, K., Madhavi, G., Rupa Devi, B., Kora, P.: Detection of Pneumonia using deep transfer learning architectures. Int. J. Adv. Trends Comput. Sci. Eng. **9**(5), 8934–8937 (2020). ISSN 2278-3091
15. Abbagalla, S., Rupa Devi, B., Anjaiah, P., Reddy Madhavi, K.: Analysis of COVID-19-impacted zone using machine learning algorithms. In: Reddy, K.A., Devi, B.R., George, B., Raju, K.S. (eds.) Data Engineering and Communication Technology. LNDECT, vol. 63, pp. 621–627. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0081-4_62
16. Prabhakar, T., Srujan Raju, K., Reddy Madhavi, K.: Support vector machine classification of remote sensing images with the wavelet-based statistical features. In: Satapathy, S.C., Bhateja, V., Favorskaya, M.N., Adilakshmi, T. (eds.) Fifth International Conference on Smart Computing and Informatics (SCI 2021), Smart Intelligent Computing and Applications, Volume 2. Smart Innovation, Systems and Technologies, vol. 283. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-9705-0_59
17. Madhavi, R., Kora, P., Reddy, L., Jangaraj, A., Soujanya, K., Prabhakar, T.: Cardiac arrhythmia detection using dual-tree wavelet transform and convolutional neural network. Soft. Comput. **26**, 3561–3571 (2022). https://doi.org/10.1007/s00500-021-06653-w
18. Swaraja, K., et al.: Brain tumor classification of MRI images using deep convolutional neural network. Traitement du Signal **38**, 1171–1179 (2021). https://doi.org/10.18280/ts.380428
19. Avanija, J., Sunitha, Gurram, Reddy Madhavi, K., Hitesh Sai Vittal, R.: An automated approach for detection of intracranial haemorrhage using DenseNets. In: Reddy, K.A., Devi, B.R., George, B., Raju, K.S. (eds.) Data Engineering and Communication Technology. LNDECT, vol. 63, pp. 611–619. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0081-4_61
20. Reddy Madhavi, K., Vijaya Sambhavi, Y., Sudhakara, M., Srujan Raju, K.: COVID-19 isolation monitoring system. In: Reddy, K.A., Devi, B.R., George, B., Raju, K.S. (eds.) Data Engineering and Communication Technology. LNDECT, vol. 63, pp. 601–609. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0081-4_60
21. Prabhakar, T., Sunitha, G., Madhavi, G., Avanija, J., Madhavi, K.R.: Automatic detection of diabetic retinopathy in retinal images: a study of recent advances. Ann. Rom. Soc. Cell Biol. **25**, 15277–15289 (2021)
22. Bhogaraju, S.D., Rajesh Kumar, K.V., Anjaiah, P., Shaik, J.H., Reddy Madhavi, K.: Advanced predictive analytics for control of industrial automation process (2021)
23. Seeja, G., Obulakonda Reddy, R., Kumar, K.V., Mounika, S.S., Reddy Madhavi, K.: Internet of Things and robotic applications in the industrial automation process. In: Goundar, S., Avanija, J., Sunitha, G., Madhavi, K., Bhushan, S. (ed.) Innovations in the Industrial Internet of Things (IIoT) and Smart Factory, pp. 50–64. IGI Global (2021). https://doi.org/10.4018/978-1-7998-3375-8.ch004
24. Avanija, J., Sunitha, G., Vittal, R.H.S.: Dengue outbreak prediction using regression model in Chittoor District, Andhra Pradesh, India. Int. J. Recent Technol. Eng. **8**(4), 10057–10060 (2019). https://doi.org/10.35940/ijrte.d9519.118419
25. Reddy, G.R.S., Rangaswamy, K., Sudhakara, M., Anjaiah, P., Madhavi, K.R.: Towards the protection and security in fog computing for industrial Internet of Things. In: Innovations in the Industrial Internet of Things (IIoT) and Smart Factory, pp. 17–32 (2021)

Artificial Intelligence for Detecting Prevalence of Indolent Mastocytosis

V. Srilakshmi[1], K. S. Chakradhar[2], K. Suneetha[3](✉), C. Shoba Bindu[4], Nagendar Yamsani[5], and K. Reddy Madhavi[6]

[1] CSE (AI and ML), B V Raju Institute of Technology, Narsapur, India
srilakshmi.v@bvrit.ac.in
[2] ECE, Mohan Babu University, Tirupati, India
[3] CS and IT, Jain (Deemed-to-be University), Bangalore, India
keerthisuni.k@gmail.com
[4] CSE, JNTUA, Ananthapuramu, India
[5] School of Computer Science and Artificial Intelligence, SR University, Warangal, India
[6] CSE, Mohan Babu University, Tirupati, India

Abstract. For the detection and elucidation of indolent mastocytosis, deep learning methods like convolution neural networks, encoder-decoders, recurrent neural networks and LSTM based frameworks are available. For studying the spread and development of mast cells causing indolent mastocytosis vast number of microscopic-imagery, micro-circulations shall be studied which is possible by artificial intelligence frameworks and neural networks. Systemic mastocytosis (SM) and related clonal mast cell disorders are underestimated as they have scarce epidemiology history. Amongst most of the subjects affected by mastocytosis, 91% are traced as the variant of indolent SM, 54.8% are traced with characteristics of bone marrow related mastocytosis. In this study, we propose a novel framework to classify the images of mastocytosis and chart out the collected, synthesized and evaluated publications and evidences related to the detection and diagnosis of indolent mastocytosis.

Keywords: systemic mastocytosis · prevalence · incidence · deep learning

1 Introduction

Mastocytosis is affect to the most common organ of the human body that has widespread complex tissues, particularly in adults. In most cases this is misidentified as bone marrow problem. W.H.O. has certified these deviations in the human beings as Urticaria Pigmentosa (UP), Solitary Mastocytoma (SM) and Diffuse Cutaneous Mastocytosis (DCM) shown in Fig. 1. The Urticaria Pigmentosa is also referred as maculopapular cutaneous mastocytosis (MPCM). The in vivo microscopy aided with epiluminescence combined with dermoscopy is applied for evaluation of microstructures and colors that belong to epidermis. The properties prevail abundantly in epidermis, papillary dermis and dermo-epidermal junctions with a widespread and ease, even they are not physically identifiable

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 33–43, 2023.
https://doi.org/10.1007/978-3-031-27524-1_4

and touchable with naked eye. One in ten thousand are affected with mastocytosis, except genetically inherited, a rare disorder, statistics are comptrolled by National Organization of Rare Disorders [1, 11] and National Institute of Health's Genetic and Rare Diseases Information Center.

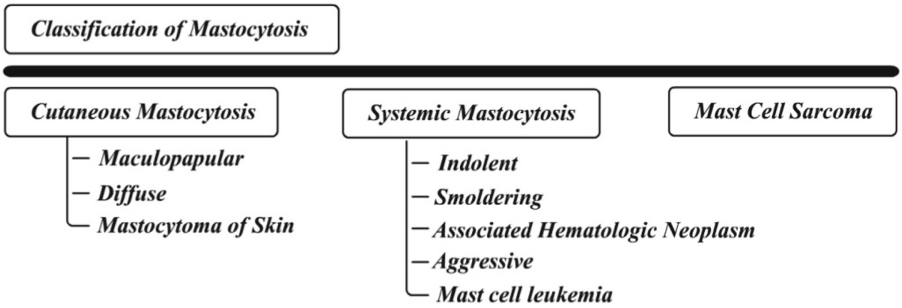


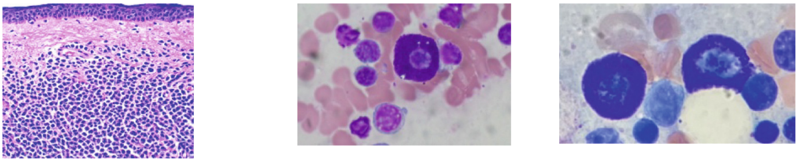
Fig. 1. WHO (2016) updated mastocytosis classification.

Mastocytosis is considered as genetic immune disorder, where the root causes and symptoms that exhibit about the occurrence of mast cell disorder are not uniform. Mast cells are immune cells that are being developed under the skin, bones and other organs as lesions and grow into lumps with complex tissues. Itchy bumpy skin with bone pains and gastrointestinal disorders are the visible symptoms of mastocytosis. A peculiar kind of white blood cells that are generally present in the human body influenced by histamine during foreign intrusion causes anaphylaxis, a painful allergy, causing itches release of mast cells as fluid from the extraneous lumps is a severe level of mastocytosis. Mastocytosis is widely classified by its eruption as cutaneous and systemic. External influence of allergies due to rapid raise and buildup of mast cells under the skin causing red or brown types of lesions is children particularly, is cutaneous. Incrementally affecting various portions of skin with the acquisition of mast cells with bone marrow, internal parts of intestines and painful allergies causing itching and effusion of fluids is systemic. A heterogeneous group of disorders characterized by increased number and accumulations of neoplastic substances causing indolence of immunity. The indolent systemic mastocytosis is 90% found in subjects affected by systemic mastocytosis.

2 Related Work

Electronically elucidating pathological characteristics, digital pathology compared to histopathology, interprets with visual presentations that contain cellular images of new sites. The advent of digital pathology paves development of methods that preprocess, transform and refine the images suitable for the analyses. The glass stains from the vet lab and the image frames from the videos are super imposed for the comparative studies and opportunities to probe into details of the subject sites. Metrics and measures for quantification of the results are easier and empirically feasible to apply compared to the vet lab records.

The reason that focuses the cause of the peculiarity in the skin tissue disorder is Ultra Violet radiations, along with UV radiations the immune systems consideration of inputs into the body. The results drawn by an expert medical analyst with visual examination and the thorough clinical diagnosis [9] will be found as deceptive in nature, due to the peculiarity of the complex tissues. The Indolent Systemic Mastocytosis requires non-invasive diagnostic methods that are interconnected with clinical methods to trace the morphological changes of lesion characteristics, are not possible to perceive with the naked eye. The morphological details are expounded by solar scanning, microscopy of epiluminescence, cross-polarization epiluminescence and side trans-illumination methods. The following are the images of narrowband UV phototherapy of systemic and indolent systemic mastocytosis.



(a) Cutaneous Mastocytosis of bone marrow (b) Histopathological Images of Indolent Systemic Mastocytosis

Fig. 2. Classification of mastocytosis

NORD is the organization for rare diseases that evolve due to abnormal symptoms that develop from time to time in the human life. The National Organization of Rare Diseases has a good amount of repositories about mastocytosis. Valent et al. ahs proposed a version of repositories for mastocytosis from the clinical observations of pediatric and non-pediatric subjects for clonal mast cell disease, otherwise mastocytosis. The classification of mastocytosis in Fig. 2, is drawn into cutaneous and systemic. Later, studies on mastocytosis have defined cutaneous mastocytosis, indolent mastocytosis, systemic mastocytosis and aggressive systemic mastocytosis. Indolent mastocytosis prevails in the mast cell leukemia and mast cell sarcoma which are variants of systemic mastocytosis. Various methods for detection are shown in Table 1. Datasets for mastocytosis are very rare and very less in the repositories. It is very difficult to maintain the repository for the types of mastocytosis, as they evolve differently from time to time and the diagnostic tests to reveal the disease is also different from one to another category of mastocytosis and particular from subject to subject. The culture tests, biopsy information of the tests belonging to mastocytosis are very vast in number and they are different from each other in the clinicians' perspectives, therefore it is very difficult to maintain the repository for the mastocytosis.

Detection of COVID-19 using Deep Learning methods has been discussed in [16], EEG-Based Brain-Electric Activity Detection During Meditation Using Spectral Estimation Techniques were found in [17], methods for Quality Improvement of Retinal Optical Coherence Tomography shown in [18], Detection of Pneumonia Using Deep Transfer Learning architectures in [19] and Analysis of COVID-19-Impacted Zone Using Machine Learning Algorithms was presented in [20]. Support Vector Machine Classification of Remote Sensing Images with the Wavelet-based Statistical Features

Table 1. Methods for detecting systemic, indolen systemic and cutaneous mastocytosis a select survey

S. no.	Authors	Year	Work
1	Chaerkady, et al. [7]	2021	The most common health problem related to skin is addressed, with the characterization of citrullination activation of mast cells is studied in this article ***Cell Characteristics and Methods:*** Analysis through Citrullinated proteins; Activation of Citrullination sites with Neutrophils
2	Massimo Salvi et al. [16]	2021	Detection of Round Cell Tumors as cutaneous and subcutaneous lesions almost as visceral anatomical locations are discussed. Majority categories of tumors as round cell tumors are discussed with histopathological analysis ***Cell Characteristics and Methods:*** Round Cell Tumors, cutaneous and subcutaneous lesions; ARCTA algorithm
3	Fatma Jendoubi et al. [10]	2021	The study in the article has projected the insights on some important symptoms of mastocytosis and the impact on patients' lives. The results are interpreted using 139 patients' data where mastocytosis is classified as cutaneous and systemic ***Cell Characteristics and Methods:*** Hamilton Score on Epidemological Data sets
4	Bruno C. Greg´orio da Silva [12]	2021	Cell detection is the most important activity in the biology. Automatic detection of cell for immunological diagnosis. Intravital video microscopy data sets are used in experiment ***Cell Characteristics and Methods:*** General Forms of Leukocytes; Deep Learning and Multi-Template Matching
5	Bruno C. Greg´orio da Silva [13]	2021	A crucial leukocyte analysis using IVM, and DCNN for IVM frame image analysis is used ***Cell Characteristics and Methods:*** Cells of Intraital Video Microscopy; Deep Convolutional Neural Networks

(*continued*)

Table 1. (*continued*)

S. no.	Authors	Year	Work
6	Leguit, Roos, et al. [6]	2020	Cases of mastocytosis are indolent and systemic are in particular. A systematic review on systemic mastocytosis and aggressive mastocytosis ***Cell Characteristics and Methods:*** Indolent, Smoldering and Aggressive. Myelodysplastic syndrome; KIT D 816 V Mutation
7	Aubreville et al. [5]	2018	The count of mitotic instances in the histopathological outcomes set as the metric for grading tumors. An Image containing Region of Interest is treated with CNN to study the mitotic instances in the images and determine tumors ***Cell Characteristics and Methods:*** Mitotic activitiy estimation, Valid mask estimation; U-Net for detecting Mast Cells
8	Gaobo Liang et al. [14]	2018	Applications of deep convolutional neural networks in study of blood-vessels and blood-related diseases is studied ***Cell Characteristics and Methods:*** Sophisticated methods of RNN and LSTM are applied on the types of blood cells with select heterogeneous frameworks
9	Peter Valent et al. [11]	2017	Neoplastic untoward developments around the tissues makeup in various areas of body with masts. Classification of systemic and cutaneous mastocytosis is discussed ***Cell Characteristics and Methods:*** Characterized study of types of mastocytosis as CM and SM, using systematic pathogenesis and diagnostic evaluation

(*continued*)

Table 1. (*continued*)

S. no.	Authors	Year	Work
10	Bains et al. [8]	2010	Clinical manifestations of mastocytosis and their therapeutic applications ***Cell Characteristics and Methods:*** A heterogeneous disorder of cells in the unprecedented organs and typically related to skin and at pediatric

was presented in [21]. Various approaches of deep learning for image classification and detection in healthcare has been presented in [22, 23, 24, 25].

2.1 Data Sets

Mast cell in human eye affected by conjunctivitis, the confocal micrograph, the image displays one mast cell causing invasion on conjuctival tissue emulated with response from inflammatory pathogen. The red dots are noted to be as mast cell histamines, which are the first cells of human immune system that react to invading pathogens. Facilitation of movement for leukocytes and other immune cells on to the infection site of the eye as represented in Fig. 3. A 2048 x 2048 image file, consisting of recorded image of confocal microscopy of mast cell with a primary cellular component of histamine of the human eye suffered with acute conjunctivitis.

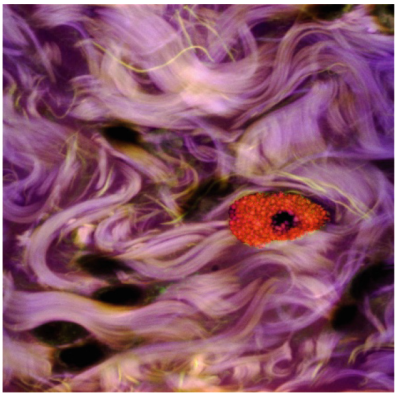


Fig. 3. Image representing projection of leukocytes emulated by histamine affected by pathogen of conjunctivitis infected human eye. This image is licensed under a creative commons attribution.

Further, annotations on data sets are given by Padanani et al. in their works: Pardanani A. "*Systemic mastocytosis: evolving lessons from large patient registry datasets*". Am J Hematol. 2016 Jul; 91(7):654-5. https://doi.org/10.1002/ajh.24395. Epub 2016 May 11. PMID: 27102564.

2.2 Methods

Microscopic medical imaging is well classified by deep learning and machine learning algorithms with transformations and augmentations. A trained histo-pathologist is an expert, who can achieve high performance and diagnostic utility with the vast knowledge of deep learning used for classification, segmentation, and labeling [2].

Table 2. Convolution, pooling and fully connected layer configuration of CNN for detection of indolent systemic mastocytosis.

Round	Layer	Kernel shape	Kernel #	Stride
1	Convolution	192 × 192 × 3	30	1
	Pooling	96 × 96 × 3	30	2
2	Convolution	96 × 96 × 3	60	1
	Pooling	48 × 48 × 3	60	2
3	Convolution	48 × 48 × 3	120	1
	Pooling	24 × 24 × 3	120	2
4	Convolution	24 × 24 × 3	240	1
	Pooling	12 × 12 × 3	240	2
5	Convolution	12 × 12 × 3	240	1
	Pooling	6 × 6 × 3	240	2

The convolution neural network deployed for the classification of micro images of mastocytosis into features of indolent systemic mastocytosis and system mastocytosis contains filters and pooling mechanism. The filters of the CNN use ReLU [14, 15], which derives various types of observations drawn from the number of iterations considering the microimage inputs. The accuracy level obtained from the inputs and the number of iterations is demonstrated as epoch graphs in Fig. 4. Epoch graphs represent the measures of adolescence of the convolutions during learning, where loss and accuracy obtained while learning with training and test data sets.

The stated CNN model consists of segmented micro images of mastocytosis of size 240 x 240 pixels, which are input to the network and an alternative convolution and pooling layers with activations using simple ReLU. Network layers shown in Table 2.The model is proven as flexible to implement in Python using Tensorflow and Keras. Overall 100 micro images are synthesized, where 70 were used in training with 40 as healthy and 30 as defective micro images. First 30 micro images are used out of which 20 as healthy and 10 as defective micro images. Training with images is iteratively processed with tensors in the convolution network. A learning rate between 0.0003 and 0.00005 is achieved for each image in the iteration, computing to 250 iterations. Mean squared error is applied to detect the error in the similarities during the experiment using synthetic micro images. As number of iterations increase the error rate is brought close to 0, thereby eliminating the probable errors in determining the similarity of images.

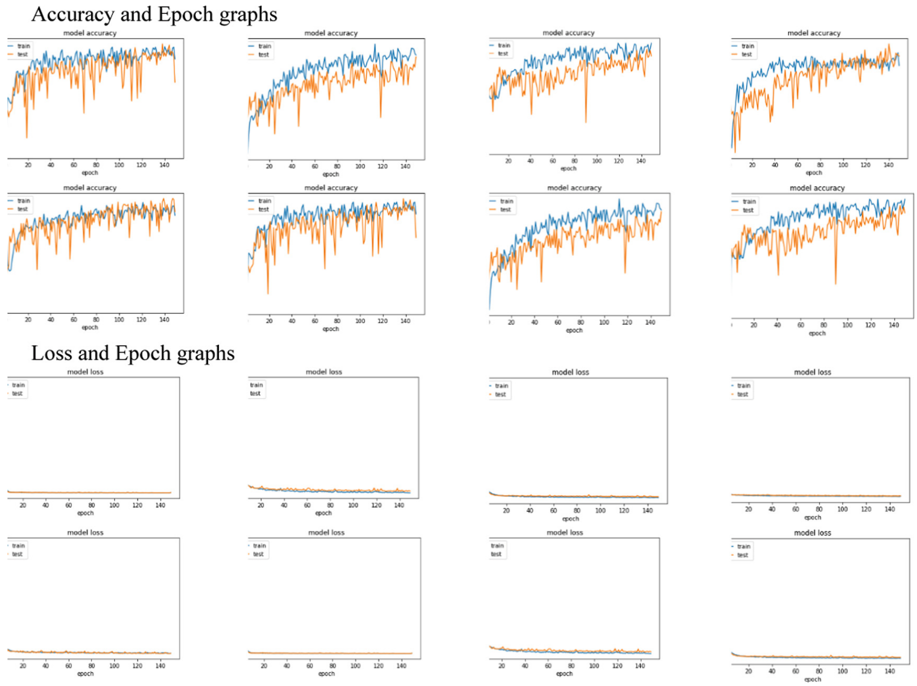


Fig. 4. The epoch graph demonstrating loss and accuracy in determining the classes of indolent systemic mastocytosis – evaluating the model of the CNN.

The deep learning AI algorithms using sequential convolutional neural network has potential of classifying the micro images of mastocytosis as system, indolent systemic. Inaccuracies are graded among the images by identifying the correct image site is formed in the micro images. The samples considered for training and testing the model is ensured with all correct sites of interest among the set of micro images.

The classification of micro images for mastocytosis has been achieved through the experiment setup with sequential convolution neural network and has extracted the categories as systemic and indolent systemic mastocytosis. While detecting the micro images with mastocytosis the proposed sequential CNN has achieved the accuracy and has been stated in the repeater-operating characteristic curves.

The data in the Table 3 demonstrates the observation of the experiment and the cumulative count of the micro images being tested for mastocytosis. The ROC curve is drawn for the said data and the Area Under the Curve is achieved as 0.889516, which is almost 88% of the images tenant the characteristics of indolent systemic mastocytosis.

The sequential model of CNN is experiment with elegant metrics where the micro images of mastocytosis are preprocessed, transformed and scaled to uniform dimensions and applied for detecting the indolence of systemic mastocytosis. The experiment is conducted on the Anaconda Navigator 3 and Jupyter with 2 Gb GPU and parallelly with Google Colab on 13 GB NVIDIA Tesla K80 GPU.

Table 3. Observations of the sequential CNN for mastocytosis in micro images.

	Recognized		*Cumulative*				
No. of Samples of Micro Images	*Detected Systemic*	*Detected Indolent Systemic*	*Correct*	*Incorrect*	*FPR*	*TPR*	*AUC*
			0	*0*	*1*	*1*	*0.082437*
50	46	3	46	3	0.9175627	0.9910979	*0.133212*
100	75	7	121	10	0.7831541	0.9703264	*0.168677*
150	97	11	218	21	0.609319	0.9376855	*0.171405*
200	102	25	320	46	0.4265233	0.8635015	*0.179509*
250	116	42	436	88	0.218638	0.7388724	*0.128442*
300	97	65	533	153	0.0448029	0.5459941	*0.010763*
350	11	82	544	235	0.0250896	0.3026706	*0.004882*
400	9	48	553	283	0.0089606	0.1602374	*0.001436*
450	5	36	558	319	0	0.0534125	*0*
500	0	18	558	337	0	0	*0*
	558	337					

3 Conclusions

The framework proposed in the experiment is a classical model of sequential CNN to classify micro images. The network is trained with sufficient size of samples where the size of samples is finalized on number of iterations. By experience the size of samples and the output of the model are fine tuned. Therefore, it is proved with a minimal vanilla model of CNN, with more economic interferences expert-pathologists can detect easily for mastocytosis with ease. Some more quality parameters and hyper parameter tuning can be setup to elevate the performance and make the framework functional on more complex micro images. As compared to any medical imaging, the micro images related to mastocytosis are varied and unique, it is very difficult to ascertain a model, still overcoming the difficulties the proposed model has achieved accuracy to the satisfactory levels and better than the experiments conducted in the wet labs.

References

1. Valent, P., et al.: The data registry of the european competence network on mastocytosis (ECNM): set up, projects, and perspectives. J. Allergy Clin. Immunol. Pract. **7**(1), 81–87 (2019)
2. Manera, M., et al.: RGB stacks gray-level analysis to study mast cell degranulation to intestinal contraction correlation in trout. J. Morphol. Sci. **30**(3) (2017)
3. Boddaert, N., et al.: Neuroimaging evidence of brain abnormalities in mastocytosis. Transl. Psychiatry **7**(8), e1197–e1197 (2017)

4. Aubreville, M., et al.: Field of interest proposal for augmented mitotic cell count: comparison of two convolutional networks. arXiv preprint arXiv:1810.09197
5. Leguit, R., et al.: The spectrum of aggressive mastocytosis: a workshop report and literature review. Pathobiology **87**(1), 2–19 (2020)
6. Pizzi, M., et al.: Crossing the borders: an integrated approach to myeloproliferative neoplasms and mastocytoses, 1492 (2021)
7. Jendoubi, F., et al.: Neuropsychiatric, cognitive and sexual impairment in mastocytosis patients. Orphanet J. Rare Diseases **16**, 1–7 (2021)
8. Valent, P., et al.: Advances in the classification and treatment of mastocytosis: current status and outlook toward the future. Cancer Res. **77**(6), 1261–1270 (2017)
9. da Silva, B.C.G., Ferrari, R.J.: Exploring deep convolutional neural networks as feature extractors for cell detection. In: Gervasi, O., Murgante, B., Misra, S., Garau, C., Blečić, I., Taniar, D., Apduhan, B.O., Rocha, A.M.A.C., Tarantino, E., Torre, C.M., Karaca, Y. (eds.) ICCSA 2020. LNCS, vol. 12250, pp. 91–103. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58802-1_7
10. da Silva, B.C.G., et al.: Detecting cells in intravital video microscopy using a deep convolutional neural network. Comput. Biol. Med. **129**, 104133 (2021). Elsevier
11. Liang, G., et al.: Combining convolutional neural network with recursive neural network for blood cell image classification. IEEE Open Access Journal with a Special Section on Trends, Perspectives and Prospects of Machine Learning Applied to Biomedical Systems in Internet of Medical Things **6**, 36188–36197 (2018)
12. Vano-Galvan, S., et al.: Dermoscopic features of skin lesions in patients with mastocytosis. Arch. Dermatol. **147**(8), 932–940 (2011)
13. Salvi, M., et al.: Histopathological classification of canine cutaneous round cell tumors using deep learning: a multi-center study. Frontiers in Veterinary Science **8**, 640944 (2021)
14. Sriharsha, A.V.: Information retrieval and linguistic analysis on cursive characters. Int. J. Mech. Eng. **7**(1), 7109–7114 (2022). Kalahari Journals
15. Galety, M.G., Al-Mukhtar, F., Rofoo, F., Sriharsha, A.V., Maaroof, R.: Electroencephalography image classification using convolutional neural networks. In: The International Conference on Innovations in Computing Research, pp. 42–52. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-14054-9_5
16. Madhavi, K.R., Madhavi, G., Krishnaveni, C.V., Kora, P.: COVID-19 detection using deep learning. In: Abraham, A., Hanne, T., Castillo, O., Gandhi, N., Rios, T.N., Hong, T.-P. (eds.) Hybrid Intelligent Systems: 20th International Conference on Hybrid Intelligent Systems (HIS 2020), December 14-16, 2020, pp. 263–269. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-73050-5_26
17. Kora, P., Rajani, A., Chinnaiah, M.C., Madhavi, K.R., Swaraja, K., Meenakshi, K.: EEG-based brain-electric activity detection during meditation using spectral estimation techniques. In: Jyothi, S., Mamatha, D.M., Zhang, Y.-D., Raju, K.S. (eds.) Proceedings of the 2nd International Conference on Computational and Bio Engineering. LNNS, vol. 215, pp. 687–693. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-1941-0_68
18. Rajani, A., Kora, P., Madhavi, K.R., Avanija, J.: Quality improvement of retinal optical coherence tomography, pp. 1–5 (2021). https://doi.org/10.1109/INCET51464.2021.9456151
19. Madhavi, K.R., Madhavi, G., Rupa Devi, B., Kora, P.: Detection of pneumonia using deep transfer learning architectures. Int. J. Adv. Trends Comput. Sci. Eng. **9**(5), ISSN 2278–3091, pp. 8934 - 8937 (2020)
20. Abbagalla, S., Rupa Devi, B., Anjaiah, P., Madhavi, K.R.: Analysis of COVID-19-impacted zone usingmachine learning algorithms. In: Ashoka Reddy, K., Rama Devi, B., Boby George, K., Raju, S. (eds.) Data Engineering and Communication Technology. LNDECT, vol. 63, pp. 621–627. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0081-4_62

21. Prabhakar, T., Srujan Raju, K., Reddy Madhavi, K.: Support Vector Machine Classification of Remote Sensing Images with the Wavelet-based Statistical Features. In: Fifth International Conference on Smart Computing and Informatics (SCI 2021), Smart Intelligent Computing and Applications, Volume 2. Smart Innovation, Systems and Technologies, vol 283. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-9705-0_59
22. Madhavi, R., Kora, P., Reddy, L., Jangaraj, A., Soujanya, K., Prabhakar, T.: Cardiac arrhythmia detection using dual-tree wavelet transform and convolutional neural network. Soft. Comput. **26**, 3561–3571 (2022). https://doi.org/10.1007/s00500-021-06653-w
23. Avanija, J., Gurram Sunitha, K., Reddy Madhavi, R., Vittal, H.S.: An automated approach for detection of intracranial haemorrhage using DenseNets. In: Ashoka Reddy, K., Rama Devi, B., Boby George, K., Raju, S. (eds.) Data Engineering and Communication Technology. LNDECT, vol. 63, pp. 611–619. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0081-4_61
24. Reddy Madhavi, K., Vijaya Sambhavi, Y., Sudhakara, M., Srujan Raju, K.: COVID-19 isolation monitoring system. In: Ashoka Reddy, K., Rama Devi, B., Boby George, K., Raju, S. (eds.) Data Engineering and Communication Technology. LNDECT, vol. 63, pp. 601–609. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0081-4_60
25. Prabhakar, T., Sunitha, G., Madhavi, G., Avanija, J., Madhavi, K.R.: Automatic detection of diabetic retinopathy in retinal images: a study of recent advances. Ann. Rom. Soc. Cell Biol. **25**, 15277–15289 (2021)

Solid Waste Management Using Deep Learning

N. S. Charan, Thumma Narasimhulu, G. Bhanu Kiran, T. Sudharshan Reddy, T. Shivangini Singh, and Gurram Sunitha(✉)

Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, A.P., India
gurramsunitha@gmail.com

Abstract. Solid waste is growing quickly as a result of urbanization, which is detrimental to human health and causes environmental damage. Waste management has often been a source of contention. The effects of improper waste management however, are inevitable. Solid waste is contributing 10% to the total global greenhouse gas emissions. The process for managing waste has not changed significantly over the last ten years, which has resulted in ineffective waste management techniques. A sophisticated waste management technology is required to maintain such a huge variety of solid waste since doing it manually would be exceedingly time-consuming and complex. Separating garbage into its many components, usually done by hand selecting, is one of the crucial tasks in waste management. We are proposing an innovative waste management classification system, constructed by exploiting the residual network ResNet-50, to streamline this procedure. It is a pretrained deep learning model that is applied to extract data from pictures and classify garbage into various categories, including plastic, glass, papers, etc. The garbage picture dataset was used to test this approach, and it demonstrated high accuracy on this dataset. By using this technique, it is simple and near instant to separate garbage without or with minimal human assistance. The ResNet-50 model has been trained with 100 epochs and the model achieved up to 82.1% of accuracy and has reached to the loss of 0.7715.

Keywords: Solid waste management · Smart systems · Classification · Image processing · Residual networks · ResNet50

1 Introduction

Waste management has often been a source of contention. The issue of waste management has frequently proven contentious. Waste management is the process of controlling waste materials from the time they are produced until they are disposed of. This deals to the way a company or business collects, transports, processes, recycles, or disposes of waste. The overall population is swelling, which has an influence on how much waste is produced. There is a vast zone for disposing of garbage all over the world, as can be shown by looking at the statistics. Major environmental concerns surround waste disposal facilities. More pollution is produced, the ozone layer is damaged, and new illnesses are transmitted. The best option, given the circumstances, could be waste management. If waste management is done correctly, we may reduce expenses and improve

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 44–51, 2023.
https://doi.org/10.1007/978-3-031-27524-1_5

efficiency in terms of the amount of garbage that is separated per interval of time. As a result, our aim is to develop a system that can effectively categories garbage from large amounts of data utilizing deep learning models and classification techniques; other steps may be incorporated in this work.

Since deep learning can analyses a huge number of characteristics for unstructured data, it has been one of the most promising technologies for classifications. Comprehensive picture recognition has been successfully accomplished using deep learning techniques. Thanks to neural networks, it is now feasible to learn about the characteristics of an image and to extract useful information from it. In the study, we present an optimal strategy for object identification that makes use of residual networks powered deep learning classification techniques.

2 Related Work

Governments are not extremely worried about the damage that might result from bad waste management, making waste management one of the areas of city development that has been neglected, especially in major cities. The effects of improper waste management however, are inevitable. The process for managing waste has not changed significantly over the last ten years, which has resulted in ineffective waste management techniques. With these techniques, the amount of waste that is recycled each day is very low. With the help of our model, we can change this by altering the process or reducing the amount of human effort required for waste management. A promising approach that has been used in numerous sectors and produced extremely precise results is deep learning and computer vision technology.

Many diverse applications have been solved by using deep learning [1–7]. Value-adding investigations have been performed and state-of-the-art technologies and automated systems have been proposed by researchers for classification of human created waste [8–14]. Convolutional neural networks have been studied for categorizing different types of electronic waste and also detect the size of the objects in the waste [8]. A multi-task CNN based learning architecture has been proven to be efficient for multi-label classification task [9].

Since image recognition is so important in practically every industry, several approaches have been created in this field; the technology may be analogous, but the practical cases for which it is applied are diverse. One well-known neural network-powered technique is YOLO, which is mostly used for real-time object recognition. Meaning of YOLO is YOLO ONLY LOOK ONCE. Object detection in YOLO, which stands for "you only look at the image once," is carried out as a regression problem and offers the class probabilities of the discovered images. For smart waste management, a YOLOv3 based approach has been efficiently generalized [10]. They used YOLO algorithm for detecting and classifying the objects in the image they capture. Residual blocks, bounding box regression, and intersection over union are the three primary components of YOLO. Each component of the YOLO has a specific purpose.

Although YOLO appears to be a great algorithm, it has a number of flaws that led us to rule it out of consideration for our problem statement. Since there are many distinct permutations of things in this highly complex picture, YOLO struggles to identify nearby objects because each grid can only suggest two bounding boxes. This reduces the likelihood that we will find objects in our dataset. Additionally, YOLO misses tiny object detection. After weighing our choices, we opted against moving forward with the YOLO. Compared to Faster RCNN, it has a lower recall and higher localization error [12].

Demographic studies have been done to identify correlations between living standards and waste management and efforts have been made to incorporate this knowledge for training the deep learning models [13]. They have considered the three types (privileged, social and poor) of residential areas in District of Al basaten of Cairo, and have performed their study in real time. The statistics of population, generation of waste, types of waste, the way of waste management by municipality were collected. Statistical analysis is done to get a basic understanding of the underlying data. Five types of solid waste from garbage of one month was collected, divided, weighed, photographed and a dataset has been created. This process is continued to collect data hundred days. The association between the living standard and the quantity of solid waste generated especially carton waste has been studied. A solid waste management prediction model was proposed using deep learning time series forecasting neural network and LSTM. It has been trained & validated and tested using data of 61 and 39 days respectively. The forecasting of the proposed model has been compared against the waste generated in the following days and has been demonstrated that the forecasting was more in line with generation of carton waste. For the remaining types of wastes, more investigation will be required. Another similar case study was executed in Shanghai [14].

Another research by Leow Wei Qin et al., highlighted the application of lightweight mobile application model designed based on MobileNetV2 [15]. The idea was that the CNN architectures are not suitable for light weight applications, especially for edge computing applications. The proposed model has been proven too be outperforming the traditional softmax and SVM models. It underperformed than the CNN, but has been proven to be comparable for light weigh applications.

With this literature review, we have acknowledged the importance of sustainable practices for waste management towards reducing carbon footprints and creating healthy and sustainable environment on earth.

3 Proposed Methodology

In search of better deep learning architectures for solid waste categorization we have investigated the suitability and efficiency of one of the residual network architectures ResNet-50 for solid waste categorization and prediction. ResNet-50 is also known as Residual Network-50. ResNet-50, which has 50 layers, was trained using a collection of 2,527 images from the dataset with 1000 different categories. There are 3.8 x 10^9 floating point operations available. On recognition tasks, ResNet-50 is renowned for its high generalization performance and low error rates. Figure 1 shows the proposed methodology for solid waste prediction using ResNet-50. Table 1 presents the configuration of neural layers in ResNet50. We propose to apply the ResNet-50 architecture for solid waste detection, categorization, and prediction for the purpose of solid waste management.

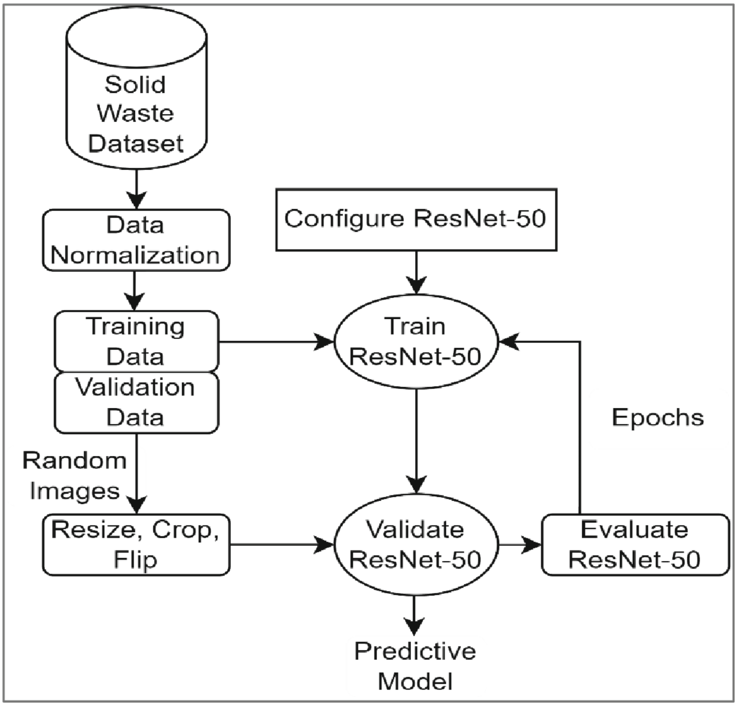


Fig. 1. The Proposed methodology for solid waste prediction using ResNet-50

Table 1. Configuration of neural layers in ResNet50.

Stages	Kernels	Layers
1	7*7, 64, stride 2	1
2	3*3 max pool, stride 2 [1*1,64 3*3,64 1*1,256]*3	9
3	[1*1,128 3*3,128 1*1,512]*4	12
4	[1*1,256 3*3,256 1*1,1024]*6	18
5	[1*1,512 3*3,512 1*1,2048]*3	9
	Average pool and soft max function	1

4 Experimentation and Results

The dataset we used for the experiment was made up of images in the.png format. Our dataset consists of a variety of trash items. These images comprise of homogeneous sorts of material; each one is debris made of glass waste, metal paper, cardboard, and plastic [16]. The dataset's sample images are displayed in Fig. 2. Figure 3 presents the bar graph illustrating how the various class types are distributed in the example data set. All the images of the dataset are normalized. The random images are selected from the validation dataset and are resized or cropped or flipped for robust validation of the learning model.

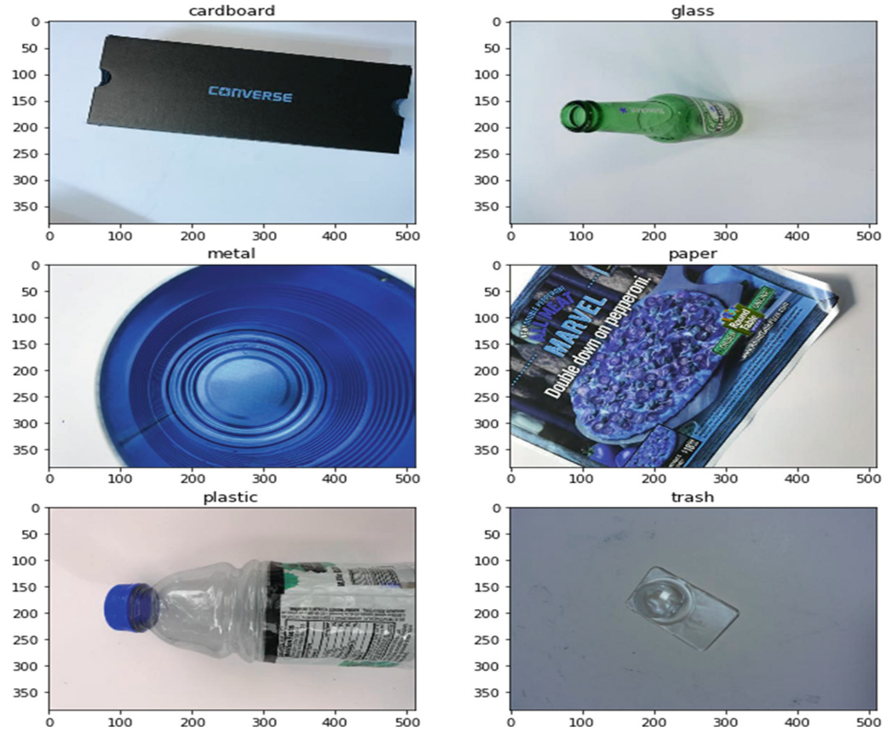


Fig. 2. Sample images of the solid waste in the dataset

For ResNet-50, the residual input is equalized in size as of the output channel. Skip connections were used to speed up the computational complexity. A downsampling layer is introduced and ReLU function is used. The default size of stride is set to 1 and for intermediate layers the stride size is set to 2. The batch normal size is set to 64 and kernel size to 7, padding to 3 for the 2-dimensional convolutional layer. For the max pool layer, the kernel size is set to 2, padding is set to 1.

Figures 4, 5 and Table 2 present the results of training the ResNet-50 for solid waste classification. Figure 4 shows the sample classified objects are labelled with according to their material ie, cardboard, plastic, paper and metal. Figure 5 shows the performance of the ResNet-50 deep neural network on training it with the solid waste dataset. Up to 150 epochs are repeated for training and validation, and Fig. 5 shows a plot of the model's accuracy. Table 2 presents the overall performance of ResNet-50 after 150 epochs of training in terms of accuracy and loss functions. The ResNet-50 model achieved upto 82.1% of accuracy and has reached to the loss of 0.7715. Overall, it can be said that ResNet-50 has perfomed well on the solid waste dataset. Fine-tuning of parameters and

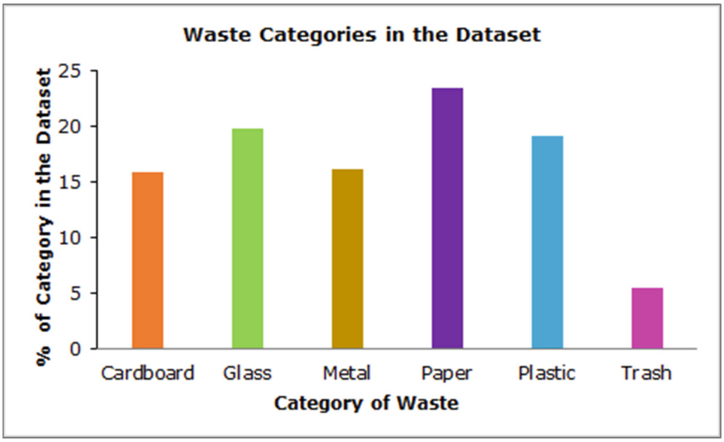


Fig. 3. Types of waste in the dataset

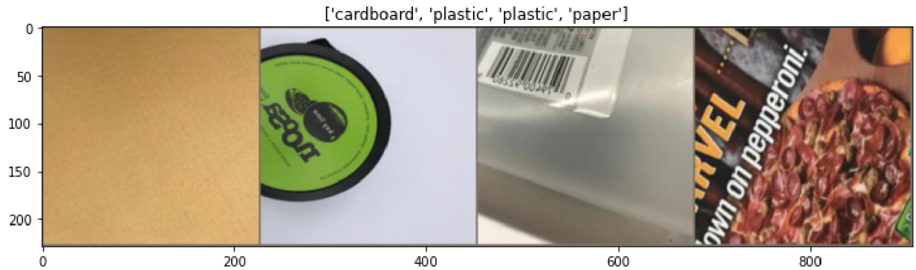


Fig. 4. Sample output of solid waste as classified by ResNet-50.

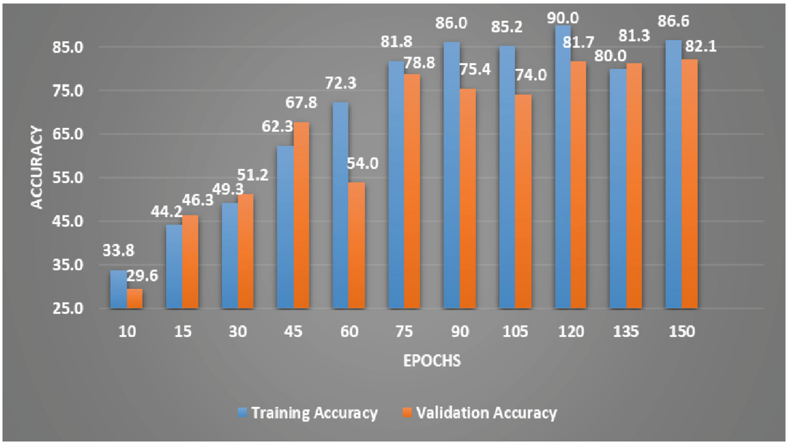


Fig. 5. Performance of ResNet-50 on solid waste dataset

pre-training of the model with large scale image datasets can be done to achieve better performance.

Table 2. Average performance of ResNet-50 on solid waste dataset

Performance metric	Average performance over number of epochs (%)
Training accuracy	90.0
Test accuracy	82.1
Training loss	77.03
Test loss	77.15

5 Conclusions

Solid waste is growing quickly as a result of urbanization, which is detrimental to human health and causes environmental damage. A sophisticated waste management system is required to maintain such a huge variety of solid waste since doing it manually would be exceedingly time-consuming and complex. Separating garbage into its many components, usually done by hand selecting, is one of the crucial tasks in waste management. We are pro-posing an innovative waste management classification system, constructed by exploiting the residual network ResNet-50, to streamline this procedure.

A pretrained deep learning model that is applied to extract data from pictures and classify garbage into various categories, including plastic, glass, papers, etc. The garbage picture dataset was used to test this approach, and it demonstrated high accuracy on this dataset. By using this technique, it is simple and near instant to separate garbage without or with minimal human assistance. The ResNet-50 model has been trained with 150 epochs and the model achieved up to 82.1% of accuracy and has reached to the loss of 0.7715. Overall, it can be said that ResNet-50 has perfomed well on the solid waste dataset. Fine-tuning of parameters and pre-training of the model with large scale image datasets can be done to achieve better performance.

References

1. Kavitha, T., et al.: Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images. Interdisc. Sci. Comput. Life Sci. **14**(1), 113–129 (2022)
2. Madhavi, K.R., Kora, P., Reddy, L.V., et al.: Cardiac arrhythmia detection using dual-tree wavelet transform and convolutional neural network. Soft. Comput. **26**(7), 3561–3571 (2022). https://doi.org/10.1007/s00500-021-06653-w
3. Reddy Madhavi, K., Sunitha, G., Avanija, J., Viswanadha Raju, S., Sindhooja, A.: Impact analysis of hydration and sleep intervention using regression techniques. Turkish J. Comput. Math. Educ. **12**(2), 2129–2133 (2021)

4. Priyanka, Y., Madhavi, K.R., Sunitha, G., Avanija, J., Meenakshi, K., Padmavathi, K.: Breast thermograms asymmetry analysis using gabor filters. In: 3rd International conference on Design and Manufacturing Aspects for Sustainable Energy, vol. 309, p. 01109 (2021)
5. Padmavathi, K., et al.: Automatic segmentation of prostate cancer using cascaded fully convolutional network. In: 3rd International conference on Design and Manufacturing Aspects for Sustainable Energy, vol. 309, p. 01068 (2021)
6. Avanija, J., et al.: Prediction of house price using XGBoost regression algorithm. Turkish J. Comput. Math. Educ. **12**(2), 2151–2155 (2021)
7. Kavitha, T., et al.: Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images. Interdisc. Sci. Comput. Life Sci. **14**, 1–17 (2021). https://doi.org/10.1007/s12539-021-00467-y
8. Nowakowski, P., Teresa, P.: Application of deep learning object classifier to improve e-waste collection planning. Waste Manage. **109**, 1–9 (2020)
9. Liang, S., Gu, Y.: A deep convolutional neural network to simultaneously localize and recognize waste types in images. Waste Manage. **126**, 247–257 (2021)
10. Kumar, S., et al.: A novel YOLOv3 algorithm-based deep learning approach for waste segregation: towards smart waste management. Electronics **10**(1), 14 (2020). https://doi.org/10.3390/electronics10010014
11. Lin, K., et al.: Toward smarter management and recovery of municipal solid waste: a critical review on deep learning approaches. J. Cleaner Production, p.130943 (2022)
12. Nowakowski, P., Pamuła, T.: Application of deep learning object classifier to improve e-waste collection planning. Waste Manage. **109**, 1–9 (2020)
13. Lin, K., Zhao, Y., Kuo, J.H.: Deep learning hybrid predictions for the amount of municipal solid waste: a case study in Shanghai. Chemosphere **307**, 136119 (2022)
14. Ahmed, A.K.A., Ibraheem, A.M., Abd-Ellah, M.K.: Forecasting of municipal solid waste multi-classification by using time-series deep learning depending on the living standard. Results in Engineering **16**, 100655 (2022)
15. Qin, L.W., et al.: Precision measurement for industry 4.0 standards towards solid waste classification through enhanced imaging sensors and deep learning model. Wireless Commun. Mob. Comput. **2021**, 1–10 (2021)
16. Dataset of images of trash. https://github.com/garythung/trashnet/blob/master/data/dataset-resized.zip
17. Wang, C., Qin, J., Qu, C., Ran, X., Liu, C., Chen, B.: A smart municipal waste management system based on deep-learning and Internet of Things. Waste Manage. **135**, 20–29 (2021)
18. Sallang, N.C.A., Islam, M.T., Islam, M.S., Arshad, H.: A CNN-based smart waste management system using tensorFlow lite and LoRa-GPS shield in Internet of Things environment. IEEE Access **9**, 153560–153574 (2021)
19. Xia, W., Jiang, Y., Chen, X., Zhao, R.: Application of machine learning algorithms in municipal solid waste management: a mini review. Waste Manage. Res. **40**(6), 609–624 (2022)
20. Neelakandan, S., et al.: Metaheuristics with deep transfer learning enabled detection and classification model for industrial waste management. Chemosphere **308**, 136046 (2022)

A Multi-stage Deep Model for Crop Variety and Disease Prediction

Rajasekhar Nennuri[1], R. Hendra Kumar[2], G. Prathyusha[3], K. Tejaswini[2], G. Kanishka[4], and Gurram Sunitha[5](✉)

[1] Department of CSE, Institute of Aeronautical Engineering, Hyderabad, Telangana, India
[2] Department of CSE, K M M Institute of Technology and Science, Tirupati, A.P, India
[3] Department of CSE, B V Raju Institute of Technology, Narsapur, Telangana, India
[4] Department of CSE, Annamacharya Institute of Technology and Sciences, Tirupati, A.P, India
[5] Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, A.P, India
gurramsunitha@gmail.com

Abstract. One of the most important and crucial professions in the world is agriculture. It is estimated that various kinds of pests like insects, weeds, animals, and diseases cause crop yield losses of 20–40%. This raises the need for detection of the types and severity of diseases at early stages, thus minimizing the usage of fertilizers and the chances of the producing healthy and higher crop yields. Therefore, it is now vital to raise the quality of agricultural products. Artificial intelligence has a great deal of potential to be demonstrated as an effective instrument that can assist agriculture in managing the growing complexity of modern agriculture. Especially doing agriculture in large scale can benefit from the intelligent systems. Till now the systems for identifying and classifying crop varieties and crop diseases have been undertaken as two different problems and the deep learning solutions developed are isolated. We have initiated and directed our work towards developing a multi-stage deep learning model which will predict the crop varieties along with the crop diseases in a composite manner. The intention was to allow integration of learning models for achieving multiple tasks. We have used ResNet18 and VGG19 for learning and classifying the crop varieties and crop diseases. We have performed experimentation with varying hyperparameter setup and the results were analyzed. The suggested model provides effectiveness by correctly detecting and recognizing crop varieties and diseases with a top-1 training accuracy of 94 .91% and a top-1 validation accuracy of 90.9%.

Keywords: Crop varieties · Crop diseases · Intelligent Systems · Modern agriculture · Multi-stage learning models

1 Introduction

One of the most important and crucial professions in the world is agriculture. Food is the basic requirement for every living being on this planet. Therefore, it is now vital to raise the quality of agricultural products. It is critical to practice proper crop management in all the stages of agricultural process. It is estimated that various kinds of pests like insects,

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 52–59, 2023.
https://doi.org/10.1007/978-3-031-27524-1_6

weeds, animals, and diseases cause crop yield losses of 20–40%. So, it is necessary to detect disease at an early stage before crop disease destroys the entire crop production. Food is the basic need essential to humans, but due to the use of a greater number of fertilizers, it increases the toxic percentage not only for plants but even for humans who are affected. This raises the need for detection of the types and severity of diseases at early stages, thus minimizing the usage of fertilizers and the chances of the producing healthy and higher crop yields. Intelligent systems have been crucial in the field of agriculture.

Agriculture is complicated and uncertain, thus intelligent systems that use customized technologies and integrated cutting-edge methodologies are needed. Agricultural resource management, field management, crop management etc. can be done effectively and economically with the help of artificial intelligence. It has a great deal of potential to be demonstrated as an effective instrument that can assist agriculture in managing the growing complexity of modern agriculture. Especially doing agriculture in large scale can benefit from the intelligent systems.

As far as our knowledge, till now the systems for identifying and classifying crop varieties and crop diseases have been undertaken as two different problems and the deep learning solutions developed are isolated. We have initiated and directed our work towards developing a multi-stage deep learning model which will predict the crop varieties along with the crop diseases in a composite manner. This will allow integration of learning models for achieving multiple tasks.

2 Literature Survey

A variety of deep learning architectures have been applied to diverse applications and have been eminent in showcasing the potential of deep learning in real-time [1–10]. The VGG19 deep learning architecture has nineteen convolution layers which are fully-connected [11]. The architecture works with zero-center normalization which is well-suited for complex functions. VGG networks work with kernel size of 3x3 thus allowing to cover each pixel in the input image along with surrounding pixels. This allows the learning model to work with improved efficiency. The authors have performed extensive and exhaustive experimentation to observe the cause effect relationships between depth of neural network and the accuracy of the neural model in cases of large-scale datasets. They have experimented the learning models with depth of sixteen and nineteen layers – VGG Net-D and VGG Net-E respectively. The ImageNet dataset contains 14 million manually labeled images along with bounding boxes. It has 1000-way classification. They have demonstrated their results on ImageNet in 2014 localization challenge. They also showcased that their results represent that their model has more generalizable using ILSVRC dataset.

Training neural networks with higher depths is a challenging task. Residual learning frameworks makes network training easier, optimizable, and accurate by making layers function to learn residual functions rather than unreferenced functions. Residual blocks are at the foundation of the deep residual networks ResNet [12]. Residual blocks are used in ResNet to boost the learning model's precision. The strength of this kind of neural network is the idea of skip connections. The skip connections are carefully customized to overcome the vanishing gradient and furthermore these models stand to take the benefit of learning the identity functions. By doing this, it is made sure that the model's later neural layers do not diminish the performance of the entire neural network by making sure that they perform at least equivalent to the earlier neural layers. The evaluation was done in comparison with VGG models by increasing the depth of the residual networks by eight times. CIFAR-10 and ImageNet datasets were used on residual networks that are even deeper than those currently being deployed. They have experimented on CIFAR-10 dataset by varying number of layers starting from 100 and up to 1000. The proposed residual neural framework won the ILSVRC 2015 challenge in classification category and COCO 2015 challenge in detection, segmentation, and localization categories.

The VGG and ResNet architectures have been researched for a variety of applications in agricultural and farming domains [13–17]. The motto of our work was to design a deep learning model which can be trained for both classification of crop varieties and crop diseases.

3 Proposed Methodology

Figure 1 shows the proposed multi-stage deep learning model for crop variety and disease prediction. Many deep learning models were proven to be efficient for predicting crop diseases. A few were proposed for predicting crop varieties. But, very limited number of models were studied for predicting composite class labels of crop variety with disease prediction. Towards this direction, our proposed multi-stage deep learning model takes as input the plant leaf image dataset and predicts a composite class label of crop variety and disease.

The proposed model works in two stages. The first stage is trained for predicting the crop variety. This stage uses the ResNet18 deep learning architecture for learning crop varieties. The models at this stage are trained using the healthy leaf images of the plants. The second stage is trained for predicting the crop disease. This stage uses the VGG19 deep learning architecture for learning crop diseases. The models at this stage are trained using both healthy and diseased leaf images of the plants. The variety of plants considered for investigation are pepper bell, potato, mango, and tomato. So, each stage in the proposed model comprises of four deep neural networks, each network being trained for one variety of plant.

The deeper the networks, the more the layers are stacked, the more accurate will be the learning model. But, the more deeper network architectures have proven to suffer with vanishing gradient. This is because of very less variation in the values of hyperparameters over number of iterations. Thus, more deeper networks can generate larger error values. The concept of residual networks is to have a shallow learning model along with a deep model as its counterpart. The shallower model transfers its knowledge to the deeper model. The deeper model is layered with shortcut connections to make network learn faster using identity mapping.

Many studies have resulted in statistics that VGG19 takes 19 billion flops and whereas ResNet18 performs with around 2 billion flops. Keeping these statistics in view, our objective was to design the model for balanced computational complexity. We have designed a two-stage network, have used shallow network (ResNet18 in the first stage and a deep network (VGG19) in the second stage.

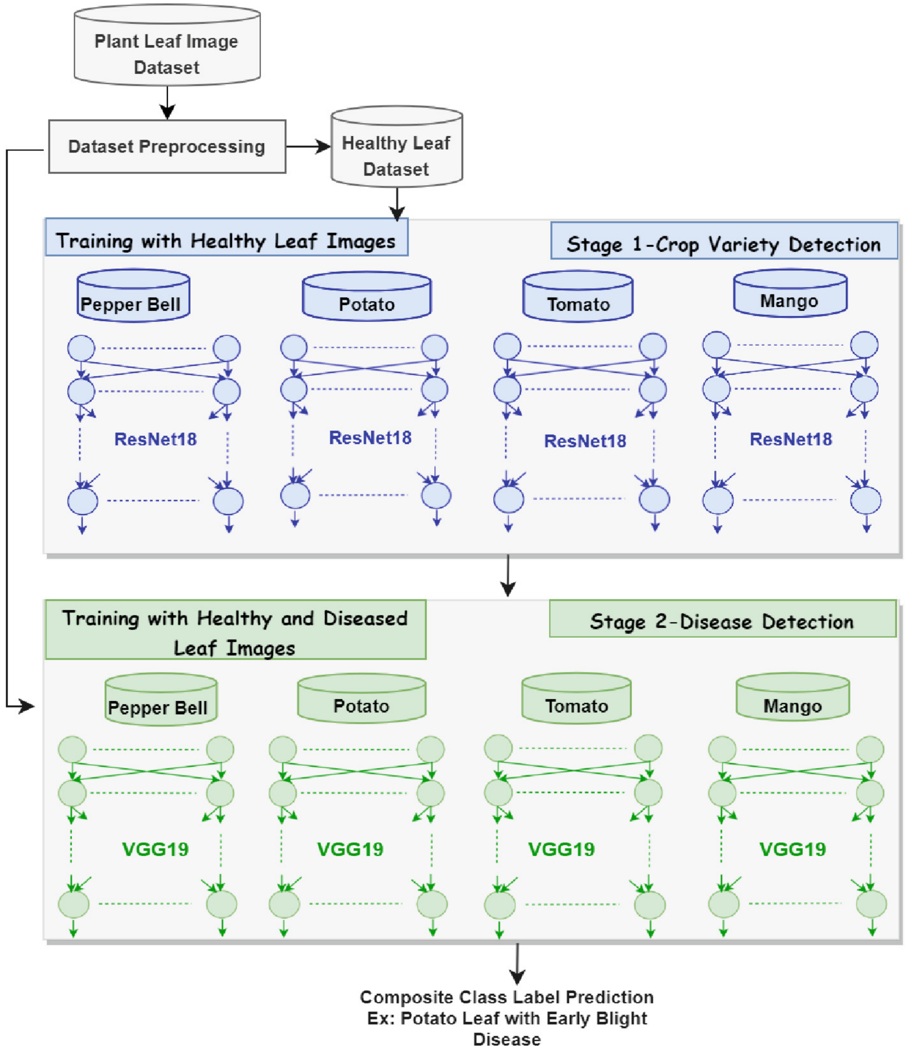


Fig. 1. The proposed multi-stage deep learning model for crop variety and disease prediction

4 Experimentation and Results

For the purpose of crop variety and disease detection, we have used the dataset comprising of plant leaf images. The variety of plants considered are pepper bell, potato, mango and tomato [18]. We have used the mango leaf dataset locally from the lebaka village, nandalur mandalam, Cuddapah district of Andhra Pradesh, India. The details of the crop varieties, leaf diseases are detailed in Table 1.

Figure 2 shows training data samples. The training data contains images of leaves with diseases at beginning stage and images where diseases have affected the plant leaves severely. As the initial step all the images are resized to 224 x 224. The images have been processed for background and noise removal. The images were also labeled. The dataset of each crop variety is skewed. Hence, as part of preprocessing of input image dataset, we have applied SMOTE for balancing the dataset.

Table 1. Details of crop varieties, diseases and dataset.

Crop variety	Leaf diseases	Number of images
Pepper bell	Healthy	997
	Bacterial spot	1478
Potato	Healthy	152
	Early blight	1000
	Late blight	1000
Tomato	Healthy	1591
	Bacterial spot	2127
	Spider mites	1676
Mango	Healthy	300
	Anthracnose	141
	Powdery mildew	133

(a) Pepper Bell Leaf with Bacterial Spot (b) Healthy Potato Leaf (c) Tomato Leaf with Spider Mites (c) Mango Leaf with Anthracnose

Fig. 2. Sample healthy and diseased leaf images of four crop varieties

The VGG19 networks were configured with kernel size of 3x3 thus allowing to cover each pixel in the input image along with surrounding pixels. This allows the learning model to work with improved efficiency. The dropout rate, learning rate, weight decay rate, batch size are kept constant at 0.5, 0.01, 0.005 and 64 respectively. Multiple runs (5) were executed to demonstrate the average performance of the proposed model. The number of epochs are varied between 50 to 150.

We have measured the performance of the proposed multi-stage model in terms of top-1 accuracy by considering the highest output value from the softmax layer as 100% confidence value. The performance statistics of the proposed model are represented in Table 2 and Fig. 3. The average top-1 accuracy of the proposed model through training

Table 2. Average performance of the proposed model for crop and disease varieties

Crop variety	Leaf diseases	Number of epochs (50–150)	
		Average top-1 training accuracy (%)	Average top-1 validation accuracy (%)
Pepper bell	Healthy	97.12	90.17
	Bacterial spot	87.44	88.76
Potato	Healthy	98.59	96.78
	Early blight	94.89	91.41
	Late blight	98.34	90.29
Tomato	Healthy	97.74	92.30
	Bacterial spot	90.59	91.45
	Spider mites	90.66	89.24
Mango	Healthy	98.20	96.88
	Anthracnose	95.38	87.29
	Powdery mildew	95.09	86.16

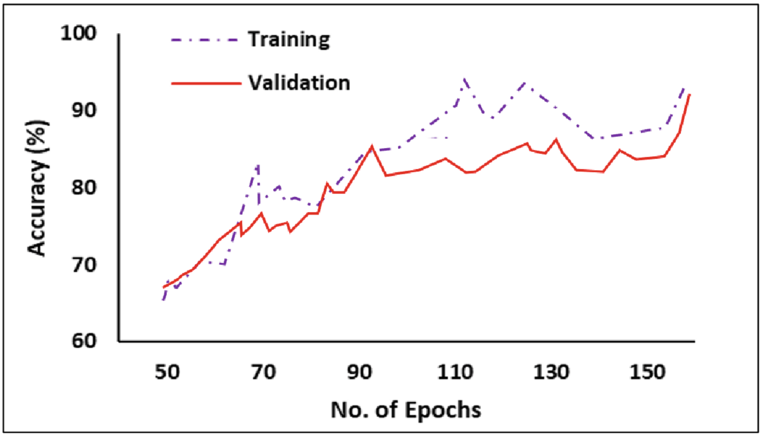


Fig. 3. Performance of multi-stage deep learning model for crop variety and disease prediction

and validation phases over number of epochs for each crop variety and disease variety are tabulated in Table 2. Figure 3 shows the performance of the proposed model through training and validation phases over number of epochs in terms of average accuracy functions. The suggested model provides effectiveness by correctly detecting and recognizing crop varieties and diseases with a top-1 training accuracy of 94 .91% and a top-1 validation accuracy of 90.9%.

5 Conclusions

Artificial intelligence has a great deal of potential to be demonstrated as an effective instrument that can assist agriculture in managing the growing complexity of modern agriculture. Especially doing agriculture in large scale can benefit from the intelligent systems. Till now the systems for identifying and classifying crop varieties and crop diseases have been undertaken as two different problems and the deep learning solutions developed are isolated. We have initiated and directed our work towards developing a multi-stage deep learning model which will predict the crop varieties along with the crop diseases in a composite manner. The intention was to allow integration of learning models for achieving multiple tasks. We have used ResNet18 and VGG19 for learning and classifying the crop varieties and crop diseases. We have performed experimentation with varying hyperparameter setup and the results were analyzed. The suggested model provides effectiveness by correctly detecting and recognizing crop varieties and diseases with a top-1 training accuracy of 94 .91% and a top-1 validation accuracy of 90.9%. Our future work intends to investigate suitability of various deep learning architectures for crop variety and crop disease detection and perform the comparative analysis. Customizing the hyperparameter setup of the deep learning architectures based on variety of crops and the variety of crop diseases is also a promising direction.

References

1. Sunitha, G., Madhavi, K.R., Avanija, J., Reddy, S.T.K., Vittal, R.H.S.: Modeling convolutional neural network for detection of plant leaf spot diseases. In: 3rd International Conference on Electronics and Sustainable Communication Systems, pp. 1187–1192. IEEE (2022)
2. Sunitha, G., Pushpalatha, M.N., et al.: Modeling of chaotic political optimizer for crop yield prediction. Intell. Autom. Soft Comput. **34**(1), 423–437 (2022)
3. Kavitha, T., Mathai, P.P., Karthikeyan, C., et al.: Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images. Interdisc. Sci. Comput. Life Sci. **14**(1), 113–129 (2022)
4. Swaraja, K., Madhavi, R., et al.: Brain tumor classification of MRI images using deep convolutional neural network. Traitement du Signal **38**(4), 1171–1179 (2021)
5. Sunitha, G., Arunachalam, R., Abd-Elnaby, M., Eid, M.M., Rashed, A.N.Z.: A comparative analysis of deep neural network architectures for the dynamic diagnosis of COVID-19 based on acoustic cough features. Int. J. Imaging Syst. Technol. **32**, 1433–1446 (2022)
6. Sunitha, G., Geetha, K., Neelakandan, S., Pundir, A.K.S., Hemalatha, S., Kumar, V.: Intelligent deep learning based ethnicity recognition and classification using facial images. Image Vis. Comput. **121**, 104404 (2022)
7. Reshma, G., et al.: Deep learning-based skin lesion diagnosis model using dermoscopic images. Intell. Autom. Soft Comput. **31**(1), 621–634 (2022)
8. Prabhakar, T., Sunitha, G., Madhavi, G., Avanija, J., Madhavi, K.R.: Automatic detection of diabetic retinopathy in retinal images: a study of recent advances. Annals of the Romanian Society for Cell Biology, pp. 15277–15289 (2021)
9. Avanija, J., Sunitha, G., Vittal, R.H.S.: Dengue outbreak prediction using regression model in Chittoor District, Andhra Pradesh, India. Int. J. Recent Technol. Eng. **8**(4), 10057–10060 (2019)

10. Avanijaa, J., et al.: Prediction of house price using XGBoost regression algorithm. Turkish J. Comput. Math. Educ. (TURCOMAT) **12**(2), 2151–2155 (2021). https://doi.org/10.17762/turcomat.v12i2.1870
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). http://arxiv.org/abs/1409.1556
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
13. Zhang, C., et al.: Resnet or densenet? Introducing dense shortcuts to resnet. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3550–3559 (2021)
14. Paymode, A.S., Malode, V.B.: Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. Artif. Intell. Agric. **6**, 23–33 (2022)
15. Saleem, M.H., Potgieter, J., Arif, K.M.: Automation in agriculture by machine and deep learning techniques: a review of recent developments. Precision Agric. **22**(6), 2053–2091 (2021)
16. Storey, G., Meng, Q., Li, B.: Leaf disease segmentation and detection in apple orchards for precise smart spraying in sustainable agriculture. Sustainability **14**(3), 1458 (2022)
17. Zhang, Y., Wa, S., Sun, P., Wang, Y.: Pear defect detection method based on ResNet and DCGAN. Information **12**(10), 397 (2021)
18. PlantVillage Dataset: Dataset of diseased plant leaf images and corresponding labels. https://www.kaggle.com/datasets/emmarex/plantdisease
19. Tugrul, B., Elfatimi, E., Eryigit, R.: Convolutional neural networks in detection of plant leaf diseases: a review. Agriculture **12**(8), 1192 (2022)
20. Yang, X., et al.: Classification of maize lodging extents using deep learning algorithms by UAV-based RGB and multispectral images. Agriculture **12**(7), 970 (2022)
21. Wongchai, A., Shukla, S.K., Ahmed, M.A., Sakthi, U., Jagdish, M.: Artificial intelligence-enabled soft sensor and internet of things for sustainable agriculture using ensemble deep learning architecture. Comput. Electr. Eng. **102**, 108128 (2022)
22. Victor, B., He, Z., Nibali, A.: A systematic review of the use of deep learning in satellite imagery for agriculture (2022). arXiv preprint arXiv:2210.01272

Arabic Machine Translation Based on the Combination of Word Embedding Techniques

Nouhaila Bensalah[1(✉)], Habib Ayad[1], Abdellah Adib[1], and Abdelhamid Ibn el Farouk[2]

[1] Team Data Science and Artificial Intelligence, University of Hassan II Casablanca, 20000 Casablanca, Morocco
nouhaila.bensalah@etu.fstm.ac.ma, abdellah.adib@fstm.ac.ma
[2] Teaching, Languages and Cultures Laboratory Mohammedia, Mohammedia, Morocco

Abstract. Automatic Machine Translation is a computer application that automatically translates one source-language sentence into the corresponding target-language sentence. With the increased volume of user-generated content on the web, textual information becomes freely available and with a gigantic quantity. Hence, it is becoming increasingly common to adopt automated analysis tools from Machine Learning (ML) to represent such kind of information. In this paper, we propose a new method called Enhanced Word Vectors (EWVs) generated using Word2vec and FastText models. These EWVs are then used for training and testing a new Deep Learning (DL) architecture based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Moreover, special preprocessing of the Arabic sentences is carried out. The performance of the proposed scheme is validated and compared with Word2vec and FastText using UN dataset. From the experimental results, we find that in most of the cases, our proposed approach achieves the best results, compared to Word2vec and FastText models alone.

Keywords: Arabic Machine Translation · Arabic preprocessing · ATB · Word embedding · Word2vec · FastText · RNN · CNN

1 Introduction

Machine Translation (MT) is one of the most sought after areas of research in the linguistics and computational community. It is a field of study under Natural Language Processing (NLP) which deals with the automatic translation of human language, from one language to another with the help of computerized systems.

Recently, Neural Machine Translation (NMT) has been expeditiously attracting the attention of the research community for its impressive results [5,23]. NMT models adhere to an end-to-end encoder-decoder approach. The encoder maps a given source sentence into a continuous vector which is termed as context vector.

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 60–69, 2023.
https://doi.org/10.1007/978-3-031-27524-1_7

This vector contains all the necessary information which can be extracted from the input sentence itself. Finally, based on the context vector, the decoder predicts its translation, word by word, in the target language. Initially, Sutskever et al. [21] proposed Recurrent Neural Networks (RNNs) for both encoding and decoding. However, one of the weaknesses of a basic RNN is that it runs into vanishing gradient due to dependencies between word pairs. In order to deal with these issues, a popular alternative is either to use LSTM [15] or GRU [10] instead of Vanilla RNN cell.

Although a lot of research has been done on NMT for European languages, few studies have used the neural approach to translate Arabic sentences [8,11,18]. Hence, one of the main purposes of this work is to combine the strengths of two widely used architectures which are: Bidirectional LSTM (BiLSTM) and Convolutional neural network (CNN) [16] and then develop a novel Deep Learning (DL) model called C-BiLSTM for Arabic MT. C-BiLSTM utilizes CNN to extract the regional features from the input sequence, and are fed into BiLSTM to deal with the long-term dependency problem and obtain the sentence representation.

Word embeddings, also called a distributed representation of words, are techniques/methods for representing an input word in such a way that it can be distinguishable from all other input words. The core idea of word embeddings is that words used in the same contexts have a high proportion of similar meaning [14]. Several methods have been developed to obtain such representations from a corpus and has achieved an important advancement in many NLP tasks [6–8].

In our work, we present the benefit obtained from the use of three different word embedding models which are: Continuous Bag of Word (CBOW) and Skip-Gram (SG) and FastText. Our scheme is based on the combination of Word2vec and FastText representations to generate EWVs for each input sentence. The resulting vectors are then used as inputs to the proposed C-BiLSTM model.

The following sections are organized as follows: The methodology is detailed in Sect. 2. The performances of the proposed method is evaluated in Sect. 3. Finally, the paper is concluded in Sect. 4.

2 Proposed Methodology

As introduced earlier, the idea of the present work is to use three word embedding models for features extraction phase. Capturing several word features from the same input sentence possibly incorporates some features those that were not captured by the other model. In this paper, after the text preprocessing step, the word representations of a given input sequence using the three embedding models are first generated. Then, the minimum (element-wise) of the resulting representations is computed using the Minimum layer to form the EWVs of the same input sequence. Finally, these word vectors are used for training and testing the proposed C-BiLSTM model. Figure 1 shows the main architecture of our work.

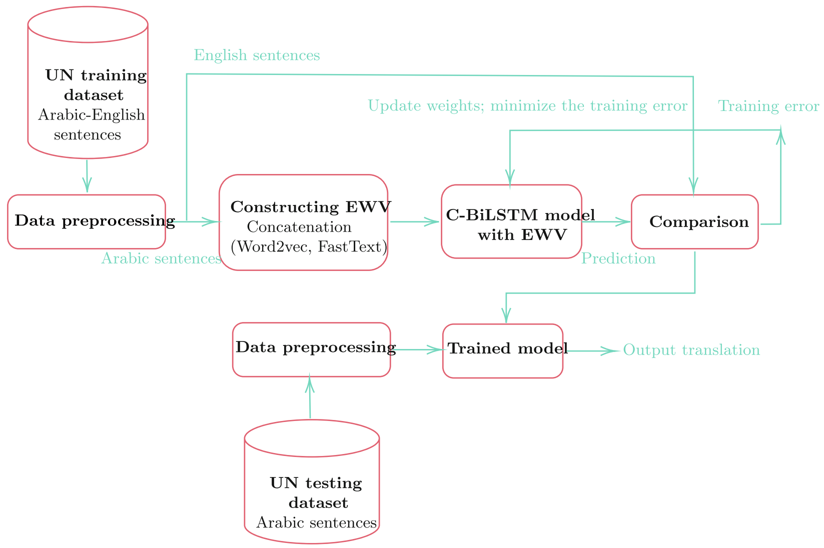


Fig. 1. The main architecture

Hereafter, we present the Arabic preprocessing used in this study, which includes the cleaning, the normalization and the tokenization of the Arabic sentences. A brief description of the different components that embody the proposed C-BiLSTM model will be given.

2.1 Arabic Preprocessing

In this paper, we aimed to investigate the performance of the proposed Arabic MT with UN dataset [22]. In order to keep only the relevant words, several preprocessing operations of Arabic sentences have been applied. First, we remove the non-Arabic letters and the tashkeel symbols. We manually correct words that have missing letters and we remove all punctuation characters. Then, we normalize some variants of Alif ا, Ya ي and Ta marbota ة in all schemes, where (آ, أ, إ) are replaced with ا, the ي is replaced with Alif Maqsura ى, the ة is replaced with ه. Finally, each Arabic sentence is transformed into a list of tokens with a morphology-aware tokenization scheme. Thereby, Farasa is used to split all clitics using SVM to rank potential word segmentations [1]. It has shown to have a significant effect on NMT [3,4,18], particularly in the case of morphologically rich languages such as Arabic.

2.2 C-BiLSTM Model

The architecture of the C-BiLSTM model is illustrated in Fig. 2. It consists of two main components: CNN and BiLSTM. In the following, we describe how we apply CNN and BiLSTM to capture both regional as well as temporal features from the EWVs features.

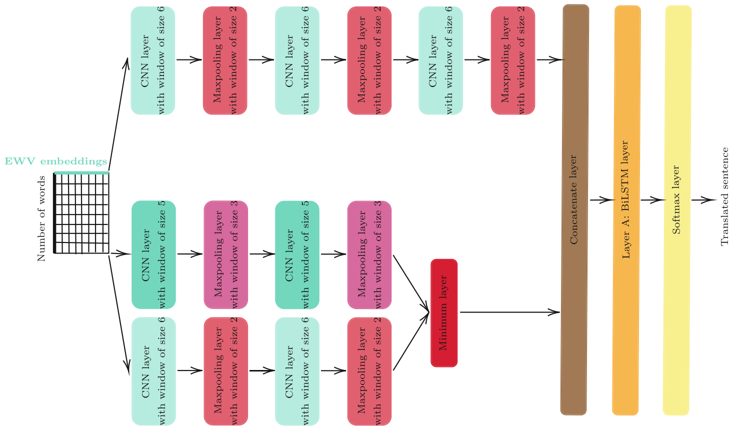


Fig. 2. The architecture of C-BiLSTM for Arabic MT

CNN. The one-dimensional convolution involves a filter (kernel) passing over a sequence and extracting features at different positions. In this work, each input sentence is modeled as a matrix by concatenating its EWV embeddings as columns. The generated matrix is used as input of three multi-layer stacked CNN networks. The first one is built by stacking three convolutional layers with a window of size six. In order to reduce the amount of parameters and computation in the network, Max pooling is added after each convolution with a pooling length of two. The second one is generated by stacking two convolutional layers with a window of size five. Max pooling is added after each convolution with a pooling length of three. The last one is built by stacking two convolutional layers with a window of size six. As the other CNN networks, Max pooling is added after each convolution with a pooling length of two. The minimum (element-wise) of the outputs generated by the second as well as the third CNN networks is computed using the Minimum layer. Finally, the outputs of the Minimum layer and those of the first CNN network are concatenated and used as inputs of the BiLSTM layer.

BiLSTM. RNNs are a type of neural networks that are capable to process sequential data $(x_1, x_2, .., x_n)$ by propagating historical information via a chain-like neural network architecture [12]. It looks at the current input x_t as well as the previous output of hidden state noted h_{t-1} at each time step to represent the sequence data. However, baseline RNNs become unable to handle long-term dependencies when the gap between the relevant information and the point where it is needed becomes large [19]. To overcome this problem, LSTM was first introduced in [15] and has risen to prominence as a state of the art in MT [21]. Then, the GRU was proposed with low complex equations. In order to make use of the amount information seen at the previous as well as the future steps, Bidirectional RNN (BiRNN) was invented [13]. Combining BiRNN with

LSTM/GRU gives Bi (LSTM/GRU) that could exploit long-range context in both input directions.

Softmax. The last layer of our proposed model is composed of a classification unit using Softmax activation function due to our multi-class problem. It gives the probability distribution of all the unique words in the target language. The predicted word at each time step is selected as the one with the highest probability.

3 Experimental Setup and Results

3.1 Hyperparameters and Training Setup

In this part, we investigated the impact of the hyperparameters on Arabic MT system to select the best ones to use during training. First, we adjust the learning rate dynamically using the LearningRateScheduler callback. At the beginning of every epoch, this callback gets the updated learning rate value from the schedule function. We run this for 100 epochs and measure the loss at each epoch starting from a learning rate of 10^{-6} to 1. We plot the results to try to find the optimal value. The optimal learning rate is approximately 2.5×10^{-3}. A manual tuning was used in order to optimize the values of the other hyperparameters. Moreover, an early stopping technique based on validation error was used.

3.2 Combination of Word Embedding Techniques

The Arabic language is known for its lexical sparsity which is due to the complex morphology of Arabic [2]. For example, the character ف (fa) in the word فكتب (fakataba) is a prefix, however, the same character in the word فرق (firaq) is an original character. To avoid this issue, we first segment the words using Farasa into stems, prefixes and suffixes. It has demonstrated to be significantly better (in terms of accuracy and speed) than the state-of-the-art segmenters; MADAMIRA [20] and Stanford [17] on MT as well as Information Retrieval (IR) tasks.

Table 1. Word vector representations training parameters

Model	Size of word vectors	Size of the context window	Number of negatives sampled	Negative
Word2Vec (CBOW/SG)	200	5	1×10^{-2}	100
FastText	200	8	2×10^{-2}	300

In order to appraise the effectiveness of the combination based word embedding techniques, investigations are conducted on a subset of the online available

UN dataset. First, we use Gensim tool[1] and Gensim's native implementation of FastText[2] to generate FastText embeddings and Word2Vec ones based on the SG and CBOW models. Table 1 shows the best hyperparameters found to generate 200 dimensional embeddings with FastText, CBOW and SG models. A random search strategy [9] is used in order to optimize the values of these parameters.

Once the word representations using the three models are generated, the next step is to select the best combination of the models' word vectors. Thereby, the minimum (element-wise), the maximum and the concatenation of the word representations of a given input sentence are computed using the Minimum, the Maximum and the Concatenate layers, respectively. The obtained results are reported in Table 2.

Table 2. Arabic MT BLEU score results

		BLEU score %
UN dataset	CBOW	57.14
	SG	56.24
	FastText	58.65
	Minimum(CBOW,SG)	58.70
	Minimum(CBOW,FastText)	57.29
	Minimum(SG,FastText)	59.20
	Minimum(CBOW, SG, FastText)	60.40
	Maximum(CBOW,SG)	58.71
	Maximum(CBOW,FastText)	59.20
	Maximum(SG,FastText)	59.46
	Maximum(CBOW, SG,FastText)	58.80
	Concatenate(CBOW,SG)	59.59
	Concatenate(CBOW, FastText)	58.41
	Concatenate(SG,FastText)	59.04
	Concatenate(CBOW, SG, FastText)	58.16

The reported results show clearly that amongst the word representations, FastText achieves better BLEU score than CBOW and SG models. This is largely due to the fact that FastText takes into consideration the internal subword information of words, which allows the model to take into account the morphology and lexical similarity of them. Further, we combined each two representations using the Concatenate layer, the Minimum layer or the Maximum one. We notice that, in the most of the cases, the translation quality is notably better than CBOW, SG and FastText models. Moreover, when combined the three word representations, we achieve the best BLEU score when the proposed DL is trained on the minimum of the three models' word vectors. With these settings, a BLEU score of 60.40 is reached compared to 57.14, 56.24 and 58.65 obtained with CBOW, SG and FastText respectively.

[1] https://radimrehurek.com/gensim/about.html.
[2] https://radimrehurek.com/gensim/models/fasttext.html.

3.3 The Impact of Different RNN Variants on Arabic MT

To study the performance of the Arabic MT under different RNNs using our model, four different variants of those are used in the layer A; see Fig. 2, which are: LSTM, GRU, BiLSTM, BiGRU.

Table 3. Arabic MT performance for different RNN variants

	BLEU score %
LSTM	57.58
GRU	57.11
BiGRU	58.50
BiLSTM	60.40

The reported results in Table 3 showed a greater gain performance of the proposed model based on BiLSTM architecture in terms of BLEU score (BLEU score = 60.40%). These findings may be explained by the fact that LSTM uses three types of gates during the training process, the input gate, the forget gate and the output one, while GRU only uses two, the update gate and the reset one. Moreover, since the computational bottleneck in our model is the softmax operation we did not remark large difference in training speed between LSTM and GRU cells. Furthermore, BiLSTM is able to exploit the historical context as well as the future one and consequently achieve best results in terms of BLEU score.

3.4 Comparison with the State-of-the-Art Works and Qualitative Evaluation

Various works have implemented a MT system based on DL. However, research has been rarely devoted to Arabic MT using a neural approach. Among these works, the authors in [4] used an encoder using two layers of BiGRU and a decoder using unidirectional GRU with the attention mechanism. They achieved an overall BLEU score of 41.14% compared to 60.40% using our approach. The authors in [18] used an encoder-decoder based on LSTM architecture with the attention mechanism. The obtained results showed that the translation quality using our approach is notably higher than that using the [18]'s model reaching a BLEU score of 60.40% compared to 42.38% obtained by the authors in [18]. This is largely due to the proposed DL (C-BiLSTM) model, the combination of the word representations as well as the morphology-based tokenization and orthographic normalization.

In Table 5, some examples of the source sentences (in Arabic) from the test set and their translations in English are illustrated. These qualitative observations demonstrates that the proposed approach translates the first three examples fluently. In the example 4, the model preserves the original meaning of the input sentence. However, in the example 5, the proposed model drops the source sentence after the comma and only translates the part before.

Table 4. Comparison with state-of-the-art works using UN dataset

	BLEU score %
[4]	41.14
[18]	42.38
Our approach	60.40

Table 5. A few examples of translations generated by our model

Source	إذ تضع في اعتبارها جميع قرارات الجمعية العامة الأخرى ذات الصلة
Our model	bearing in mind all other relevant general assembly resolutions
Truth	bearing in mind all other relevant general assembly resolutions
Source	إذ تشير إلى قراراتها السابقة بشأن التعاون بين الأمم المتحدة وجامعة الدول العربية
Our model	recalling its previous resolutions on cooperation between the united nations and the league of arab states
Truth	recalling its previous resolutions on cooperation between the united nations and the league of arab states
Source	مؤتمر الأمم المتحدة الثالث المعني بأقل البلدان نموا
Our model	third united nations conference on the least developed countries
Truth	third united nations conference on the least developed countries
Source	٢ تلاحظ مع الارتياح الدعم الذي قدمه البلد المضيف من أجل إنشاء المركز
Our model	2 notes with satisfaction the support provided for the establishment of the centre by the host country
Truth	2 notes with satisfaction the support given to the establishment of the centre by the host country
Source	الزمالات والتدريب والخدمات الاستشارية للأمم المتحدة في ميدان نزع السلاح
Our model	united nations disarmament fellowship and and
Truth	united nations disarmament fellowship, training and advisory services

4 Conclusion

In this paper, we aimed to find a robust set of features for Arabic MT. For this propose, performance analysis of a combination of three models, namely, CBOW, SG and FastText, has been discussed using UN dataset. Moreover, a DL architecture based mainly on BiLSTM and CNN has been used for the task of MT between English and Arabic texts. The obtained results were compared with those using CBOW, SG and FastText. It has been revealed that the proposed scheme exhibits a high BLEU compared to CBOW, SG, FastText as well as the sate-of-the-art works. Further studies will aim at developing a technique that allows the model to automatically search for parts of the word representations that are relevant to predicting the translation of a source sentence.

Funding. This work was supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the CNRST of Morocco (Alkhawarizmi/2020/01).

References

1. Abdelali, A., Darwish, K., Durrani, N., Mubarak, H.: Farasa: a fast and furious segmenter for Arabic. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 11–16 (2016)
2. Al-Sallab, A., Baly, R., Hajj, H., Shaban, K.B., El-Hajj, W., Badaro, G.: Aroma: a recursive deep learning model for opinion mining in Arabic as a low resource language. ACM Trans. Asian Low Resour. Lang. Inf. Process. **16**, 4 (2017)
3. Almahairi, A., Cho, K., Habash, N., Courville, A.C.: First result on Arabic neural machine translation. CoRR abs/1606.02680 (2016)
4. Alrajeh, A.: A recipe for Arabic-English neural machine translation. CoRR abs/1808.06116 (2018)
5. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations, ICLR (2015)
6. Bensalah, N., Ayad, H., Adib, A., Farouk, A.I.E.: Arabic sentiment analysis based on 1-D convolutional neural network. In: International Conference on Smart City Applications, SCA20, Safranbolu, Turkey (2020)
7. Bensalah, N., Ayad, H., Adib, A., Farouk, A.I.E.: Combining word and character embeddings in Arabic chatbots. In: Advanced Intelligent Systems for Sustainable Development, AI2SD 2020, Tangier, Morocco (2020)
8. Bensalah, N., Ayad, H., Adib, A., Farouk, A.I.E.: CRAN: an hybrid CNN-RNN attention-based model for Arabic machine translation. In: International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications, CloudTech 2020, Marrakesh, Morocco (2020)
9. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. **13**, 281–305 (2012)
10. Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: encoder–decoder approaches. In: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pp. 103–111 (2014)

11. Durrani, N., Dalvi, F., Sajjad, H., Vogel, S.: QCRI machine translation systems for IWSLT 16. CoRR abs/1701.03924 (2017)
12. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. The MIT Press (2016)
13. Graves, A., Mohamed, A., Hinton, G.E.: Speech recognition with deep recurrent neural networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, pp. 6645–6649 (2013)
14. Harris, Z.: Distributional structure. Word **10**(2–3), 146–162 (1954)
15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
16. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**, 2278–2324 (1998)
17. Monroe, W., Green, S., Manning, C.D.: Word segmentation of informal Arabic with domain adaptation. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 206–211, June 2014
18. Oudah, M., Almahairi, A., Habash, N.: The impact of preprocessing on Arabic-English statistical and neural machine translation. In: Proceedings of Machine Translation Summit XVII Volume 1: Research Track, MTSummit, pp. 214–221 (2019)
19. Pascanu, R., Mikolov, T., Bengio, Y.: On the difficulty of training recurrent neural networks. In: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, ICML 2013, JMLR.org, pp. III-1310–III-1318 (2013)
20. Pasha, A., et al.: MADAMIRA: a fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), pp. 1094–1101, May 2014
21. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, pp. 3104–3112, 8-13 December 2014
22. Tiedemann, J.: Parallel data, tools and interfaces in OPUS. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey, pp. 2214–2218 (2012)
23. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems 30, 5998–6008 (2017)

Fused Local Color Pattern (FLCP): A Novel Color Descriptor for Face Recognition

Shekhar Karanwal(✉)

Department of CSE, Graphic Era University, Dehradun, UK, India
shekhar.karanwal@gmail.com

Abstract. Literature suggests that when color information is used then there is immense improvement in accuracy. This work presented such novel descriptor so-called Fused Local Color Pattern (FLCP) by using the RGB color format. Precisely from R component MRELBP-NI is imposed for feature extraction, from G component 6×6 MB-LBP is imposed for feature extraction and from B component RD-LBP is imposed for feature extraction. The features extracted from all three components are joined which is called as FLCP. Compaction and matching is done by PCA and SVMs. Results on GT confirms the effectiveness of the FLCP descriptor as compared to the individually implemented gray scale based descriptors. FLCP also outperforms various literature methods. The novelty of proposed work is the development of discriminant descriptor by fusing the features extracted from RGB color format. This concept is lacking in the previous work.

Keywords: Gray scale based descriptors · color descriptors · feature compression · classification

1 Introduction

In last two decades local descriptors have gained huge popularity because of its discriminativity in unconstrained conditions. The local descriptors are much effective and efficient than global ones. In local, extraction of features are done from distinct image areas such as eyes, nose, mouth and forehead. The features extracted from these areas are then joined to build the complete feature size. In global, the extraction of features are done from the whole face image, which is then consumed by the classifier for the performance evaluation. The global methods are useful as feature compression. It is because it removes the redundant features for classification and select the essential features. The uncontrolled conditions possesses by image are emotion, light, pose, bur, noise, occlusion and corruption.

Most of earlier work reported in literature is based on introducing the gray scale based descriptors. There is very less work reported in literature where discriminativity is enhanced by using the color information. By using color information the discriminativity is improved as compared to the gray scale base descriptors. Karanwal et al. [1] discovered CZZBP and CMBZZBP for FR. The color component used for the feature extraction is RGB. In CZZBP, for three components the zigzag designs are created and then features

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 70–79, 2023.
https://doi.org/10.1007/978-3-031-27524-1_8

are extracted as per the zigzag designs. Then all component features are merged to develop the CZZBP size. In CMBZZBP, similar procedure is conducted on median patch. Experiments proves that both color descriptors are far effective than the gray scale based descriptors. Zhu et al. [2] developed the OC-LBP descriptor for several applications. Precisely, 3×3 patch is decomposed into two groups and then in each orthogonal group there is collation of neighbors to center pixel. Both OLBP codes are merged to build OC-LBP size. Color based OC-LBP comprehensively outperforms gray scale based OC-LBP. For extracting color OC-LBP different color formats are utilized. Results are performed on various applications. Karanwal et al. [3] invented the DCD descriptor by using color parts of HELBP, LBP and LPQ. The color parts are generated from R, G and B channels of RGB format. Further all channel features are merged to create DCD size. DCD proves effective than various others. Experiments are conducted on GT face dataset. Agarwal et al. [4] invented the MC-LTP for various applications. The color features are acquired by joining the V-V, H-V & S-V channels (of HSV space) features. Experiments confirms the potency of the developed method.

Inspiring from the literature work, the proposed work presented such novel descriptor so-called Fused Local Color Pattern (FLCP) by using the RGB color format. Precisely from R component the MRELBP-NI [5] is imposed for feature extraction, from G component the 6×6 MB-LBP [6] is imposed for feature extraction and from B component RD-LBP [7] is imposed for feature extraction. The features extracted from all three components are joined which is called as FLCP. Compaction and matching is done by PCA [8] and SVMs [9]. Results on GT [10] confirms proves the effectiveness of the FLCP descriptor as compared to the individually implemented gray scale based descriptors. FLCP also outperforms various literature methods.

Road Map: Related works are elaborated in Sect. 2, description of techniques are placed in Sect. 3, outcomes are carried out in Sect. 4 and conclusion with future directions are done Sect. 5.

2 Related Works

Chaudhari et al. [11] develops banana leaf disease classification method by using LBP and GLCM. As LBP capture local features and have good recognition & classification performance therefore used. On the other side GLCM is used for making the feature size. Experiments shows that the proposed method achieve stupendous outcomes. Karanwal et al. [12] invented the ILBP for Face Recognition (FR). LBP doesn't build balanced feature map for feature extraction therefore ILBP is proposed. In ILBP, two major steps are considered and these are, neighbor's comparison with its mean and –ve threshold is utilized for comparison. By using both these steps the transformed image produces much better histogram feature in contrast to LBP and others. The two datasets used for the evaluation is ORL and GT. Khanna et al. [13] makes use of the combination of STFT and LBP for Expression Recognition (ER). STFT is used form acquiring frequency domain features and LBP acquires local features. Further FDR, chi-square test and variance threshold are also presented for compaction. Finally SVMs is used for classification. Experiments on JAFFE dataset confirms the method efficacy. Vu et al. [14] imposed masked based FR by integrating the CNN and LBP. Initially the CNN method RetinaFace is utilized as a fast and efficient encoder, which jointly learns the self-supervised and extra-supervised features from various scales. Furthermore LBP features extracted from eyes, eyebrows, nose and forehead areas are merged with the features learnt earlier from the RetinaFace. Results reflects potency of method.

Wang et al. [15] discovered novel method for Texture Analysis (TA) by fusing the global and local features. Initially image pyramid (multiscale) space is constructed to observe the scale variations. Then Gabor features are extracted from the image pyramid space which is then used as the global feature. Then Completed Local Binary Count (CLBC) is deployed to the original image (at distinct radius) which is then used as the local feature. Ultimately global and local features are merged to develop entire feature representation. Experiments shows the developed method efficacy. Karanwal et al. [16] develops blur robust operator BILBD for FR by using three state of art descriptors and these are MRELBP-NI, RD-LBP and LPQ. MRELBP-NI captures the microstructure and macrostructure details, RD-LBP captures the radial information and LPQ captures the frequency domain features. The merits of these three descriptors are incorporated in one framework called as BILBD. Experiments on 3 different datasets confirms BILBD efficacy. Karanwal et al. [17] invented MB-ZZLBP for FR. First filtration of mean is done. Then zigzag oriented pixels are compared to develop MB-ZZLBP code. Experiments proves MB-ZZLBP efficacy on two datasets. Borlea et al. [18] provides the technique for enhancing the performance of resulted clusters produced by algorithm K-means by resulted clusters after processing in conjunction with supervised algorithm (learning).

3 Description of Descriptors

3.1 MRELBP-NI

This descriptor was proposed for TA [5]. Due to acquiring the microstructure and macrostructure details MRELBP-NI is very fruitful in various unfavorable conditions. Its detailed description is defined as: Initially median value is obtained from nine regions of the 9×9 patch and each region size is 3×3. Then all the neighbors are thresholded to label 1 for the condition of higher or equal value to mean. Else 0 is given. This leads to pattern length of eight bits, which is transfigured to MRELBP-NI code by weights allocation and summing of values. By computing MRELBP-NI code for every pixel place there is formation of MRELBP-NI image and size is 256. Equation (1) and Eq. 2 portrays MRELBP-NI description. In Eq. (1) P, R_2, $\left(\phi(W_{R_2,P,3,\,p})\right)$ and $\mu_{R_2,P,3}$ signifies neighborhood limit, radius, filter (median) and mean.

$$\mathrm{MRELBP-NI}_{P,R_2,3} = \sum\nolimits_{p=0}^{P-1} h(\phi\left(W_{R_2,P,3,\,p}\right) - \mu_{R_2,P,3})2^{p},$$

$$h(x) = \begin{pmatrix} 1 & x \geq 0 \\ 0 & x < 0 \end{pmatrix} \tag{1}$$

$$\mu_{R_2,P,3} = \frac{1}{P}\sum\nolimits_{p=0}^{P-1} \phi\left(W_{R_2,P,3,\,p}\right) \tag{2}$$

3.2 6 × 6 MB-lBP

MB-LBP [6] was introduced for FR. Due to acquiring the microstructure and macrostructure details the descriptor is very fruitful in the various unconstrained conditions. Its detailed description is defined as: In 6 × 6 MB-LBP, mean value is obtained from nine blocks of 6 × 6 patch and each region size is 2 × 2. Then all neighbors are thresholded to label 1 for bigger or similar center value. Else 0 is given. This builds pattern length of eight bits, which is transfigured to the 6 × 6 MB-LBP code by weights apportion. By computing 6 × 6 MB-LBP code for every pixel place there is development of 6 × 6 MB-LBP image and size is 256. Equation 3 and Eq. 4 shows its illustration. Equation 3 generates square regions ($L_{i,j}$) mean and in Eq. 4 P, R, $W_{R,p}$ and W_c states sized of neighborhood, radius, sole positions and center pixel.

$$W_{i,j} = \mathrm{mean}\left(L_{i,j}\right) \tag{3}$$

$$6 \times 6\ \mathrm{MB} - \mathrm{LBP}_{P,R}(x_c) = \sum\nolimits_{p=0}^{P-1} h\left(W_{R,p} - W_c\right)2^p,\ h(x) = \begin{pmatrix} 1 & x \geq 0 \\ 0 & x < 0 \end{pmatrix}, \tag{4}$$

3.3 RD-LBP

RD-LBP [7] was invented also for TA. Due to acquiring the radial information the RD-LBP prove out very effective descriptor. The detailed illustration of RD-LBP is defined as: In RD-LBP, radial differences generated among two different scales (of 5 × 5 patch) are thresholded to 1 for the condition of higher or similar value to 0. Otherwise label 0 is assigned. This leads to pattern length of eight bits, which is converted to RD-LBP code by apportion of weights. By evaluating RD-LBP code in each place (pixel) the RD-LBP image develops, which creates the 256 size. Equation 5 shows the RD-LBP description. In Eq. 5, P, R_1, R_2, $W_{R_2,p}$ and $W_{R_1,p}$ specifies size of neighborhood, scales (radius R_1 and R_2), sole pixels at R_2 and sole pixels at R_1.

$$\mathrm{RD} - \mathrm{LBP}_{P,\ R_1,R_2} = \sum\nolimits_{p=0}^{P-1} h(W_{R_2,p} - W_{R_1,p})2^p,\ h(x) = \begin{pmatrix} 1 & x \geq 0 \\ 0 & x < 0 \end{pmatrix} \tag{5}$$

3.4 Fused Local Color Pattern (FLCP)

Literature reports some of the local descriptors where discriminativity is improved by using color information in contrast to gray scale counterparts. This paper proposed such color descriptor so-called FLCP. In FLCP, initially RGB format is taken and from R component MRELBP-NI is used for feature extraction, from G component 6 × 6 MB-LBP is utilized and from B component RD-LBP is utilized. Features generated from the color channels are integrated to develop the FLCP size. Therefore FLCP builds 768 size (as each component feature forms 256 size). The compaction and matching is done by PCA and SVMs. Figure 1 shows complete FR framework.

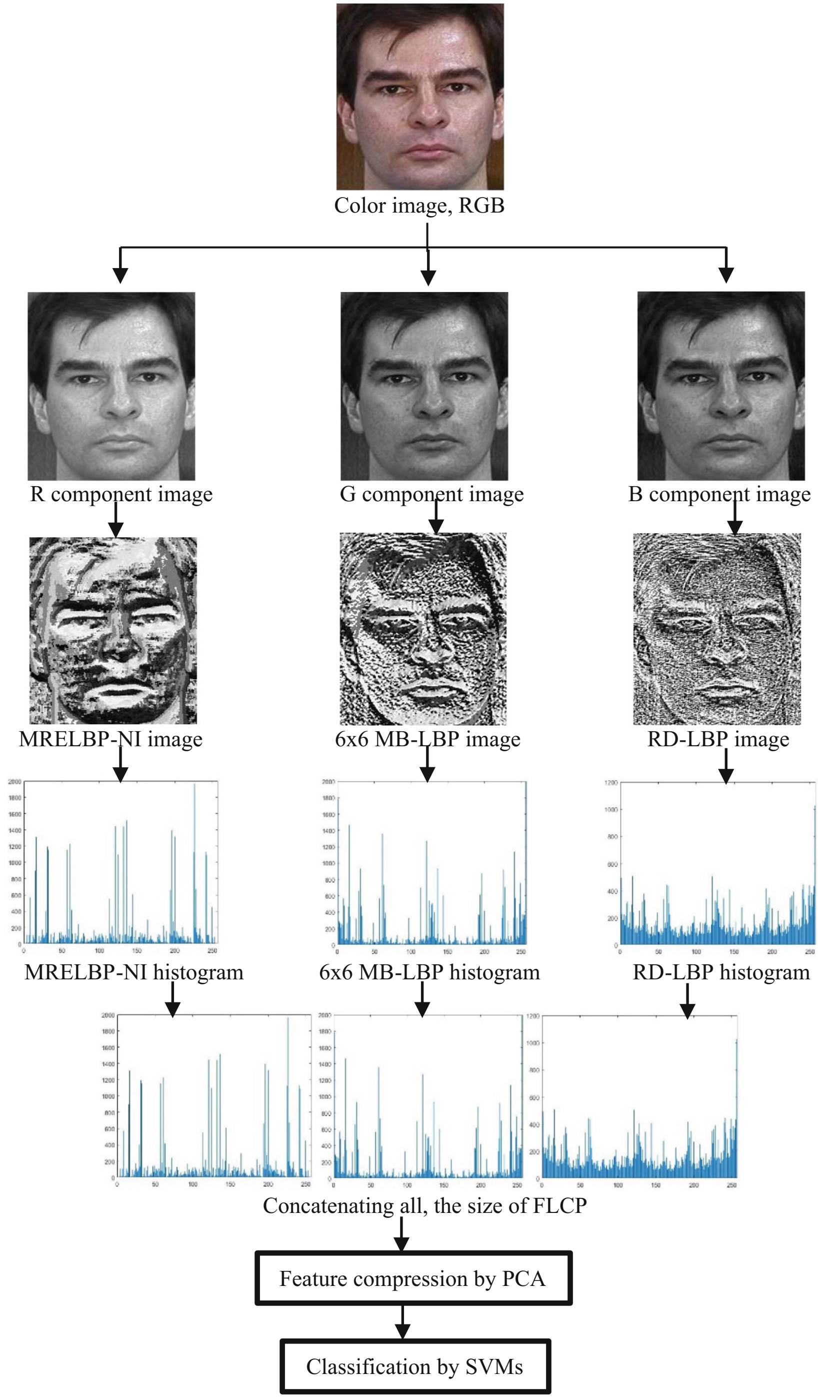


Fig. 1. The proposed FR framework

4 Experiments

4.1 Dataset Specification

GT dataset is equipped with 750 color samples of 50 individuals and each individual carry 15 images. These 15 images show the configurations (challenges) of pose, light, emotion and scale variations. As scale variations persist on images therefore image resolution is inconsistent. Some samples of the GT dataset are shown in Fig. 2.

Fig. 2. Some images of GT dataset

4.2 Feature Size Details

The color transformation to gray is performed for the compared ones and then images are downsampled to 52×48. Then MRELBP-NI, RD-LBP and 6x6 MB-LBP are utilized. All three develops the feature size of 256. For proposed descriptor initially the color image is resized to 52×48 and then it is decomposed into R, G and B components. From R component MRELBP-NI is imposed for feature extraction, from G component 6×6 MB-LBP is imposed for feature extraction and from B component RD-LBP is imposed for feature extraction. All respective component feature develops the size of 256 so FLCP size is the concatenated histograms of all three. Finally FLCP develops the 768 size. After PCA the size evolved is 25. The detailed investigation of all the parameter settings are displayed in Table 1. MATLAB R2021a is used for all testing.