

GLOBAL JOURNAL

OF COMPUTER SCIENCE AND TECHNOLOGY: D

Neural & AI

Machine Learning Solution

Human Tracking and Profiling

Highlights

Artificial Neural Networks

Classification of Kashmiri Language

Discovering Thoughts, Inventing Future

VOLUME 22 ISSUE 1 VERSION 1.0



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: D
NEURAL & ARTIFICIAL INTELLIGENCE



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: D
NEURAL & ARTIFICIAL INTELLIGENCE

VOLUME 22 ISSUE 1 (VER. 1.0)

OPEN ASSOCIATION OF RESEARCH SOCIETY

© Global Journal of Computer Science and Technology. 2022.

All rights reserved.

This is a special issue published in version 1.0 of "Global Journal of Computer Science and Technology" By Global Journals Inc.

All articles are open access articles distributed under "Global Journal of Computer Science and Technology"

Reading License, which permits restricted use. Entire contents are copyright by of "Global Journal of Computer Science and Technology" unless otherwise noted on specific articles.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission.

The opinions and statements made in this book are those of the authors concerned. Ultraculture has not verified and neither confirms nor denies any of the foregoing and no warranty or fitness is implied.

Engage with the contents herein at your own risk.

The use of this journal, and the terms and conditions for our providing information, is governed by our Disclaimer, Terms and Conditions and Privacy Policy given on our website <http://globaljournals.us/terms-and-condition/menu-id-1463/>

By referring / using / reading / any type of association / referencing this journal, this signifies and you acknowledge that you have read them and that you accept and will be bound by the terms thereof.

All information, journals, this journal, activities undertaken, materials, services and our website, terms and conditions, privacy policy, and this journal is subject to change anytime without any prior notice.

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors: Open Association of Research Society

Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America

USA Toll Free: +001-888-839-7392

USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please
email us at local@globaljournals.org

eContacts

Press Inquiries: press@globaljournals.org
Investor Inquiries: investors@globaljournals.org
Technical Support: technology@globaljournals.org
Media & Releases: media@globaljournals.org

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

EDITORIAL BOARD

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

Dr. Corina Sas

School of Computing and Communication
Lancaster University Lancaster, UK

Dr. Sotiris Kotsiantis

Ph.D. in Computer Science, Department of Mathematics,
University of Patras, Greece

Dr. Diego Gonzalez-Aguilera

Ph.D. in Photogrammetry and Computer Vision Head of
the Cartographic and Land Engineering Department
University of Salamanca Spain

Dr. Yuanyang Zhang

Ph.D. of Computer Science, B.S. of Electrical and
Computer Engineering, University of California, Santa
Barbara, United States

Dr. Osman Balci, Professor

Department of Computer Science Virginia Tech, Virginia
University Ph.D. and M.S. Syracuse University, Syracuse,
New York M.S. and B.S. Bogazici University, Istanbul,
Turkey

Dr. Kwan Min Lee

Ph. D., Communication, MA, Telecommunication,
Nanyang Technological University, Singapore

Dr. Khalid Nazim Abdul Sattar

Ph.D, B.E., M.Tech, MBA, Majmaah University,
Saudi Arabia

Dr. Jianyuan Min

Ph.D. in Computer Science, M.S. in Computer Science, B.S.
in Computer Science, Texas A&M University, United States

Dr. Kassim Mwitondi

M.Sc., PGCLT, Ph.D. Senior Lecturer Applied Statistics/
Data Mining, Sheffield Hallam University, UK

Dr. Kurt Maly

Ph.D. in Computer Networks, New York University,
Department of Computer Science Old Dominion
University, Norfolk, Virginia

Dr. Zhengyu Yang

Ph.D. in Computer Engineering, M.Sc. in
Telecommunications, B.Sc. in Communication Engineering,
Northeastern University, Boston, United States

Dr. Don. S

Ph.D in Computer, Information and Communication
Engineering, M.Tech in Computer Cognition Technology,
B.Sc in Computer Science, Konkuk University, South
Korea

Dr. Ramadan Elaiess

Ph.D in Computer and Information Science, University of
Benghazi, Libya

Dr. Omar Ahmed Abed Alzubi

Ph.D in Computer and Network Security, Al-Balqa Applied
University, Jordan

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Dr. Rajesh Kumar Rolan

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - MICROSOFT, SCJP - SUN MICROSYSTEMS, SINGHANIA UNIVERSITY, INDIA

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Pranit Gopaldas Shah

MTech CE, BECE, MPM, FCSRC, Master of Technology in Computer Engineering, Parul University, India

CONTENTS OF THE ISSUE

- i. Copyright Notice
 - ii. Editorial Board Members
 - iii. Chief Author and Dean
 - iv. Contents of the Issue
-
1. Sentiment Polarity Identification of Social Media Content using Artificial Neural Networks. *1-8*
 2. Human Tracking and Profiling for Risk Management. *9-15*
 3. Deep CNN Model for Non-Screen Content and Screen Content Image Quality Assessment. *17-24*
 4. Acoustic Features based Accent Classification of Kashmiri Language using Deep Learning. *25-29*
 5. Neural Network Design using a Virtual Reality Platform. *31-47*
 6. Vehicle Routing Problem with Time Window Constrain using KMeans Clustering to Obtain the Closest Customer. *49-64*
-
- v. Fellows
 - vi. Auxiliary Memberships
 - vii. Preferred Author Guidelines
 - viii. Index



Sentiment Polarity Identification of Social Media Content using Artificial Neural Networks

By K. Victor Rajan & Brittney Jackson

Atlantic International University

Abstract- Sentiment of people about consumer goods and government policies for decision making is normally collected through feedback forms, surveys etc. The social network sites and micro blogging sites are considered a very good source of information nowadays because people share and discuss their opinions about a certain topic freely. With the increased use of technology and social media, people proactively express their opinion through social media sites like Twitter, Facebook, Instagram etc. A social media sentiment analysis can help companies to understand how people feel about their products. On the other hand, extracting the sentiment from social media text is a challenging task due to the complexity of natural language processing of social media language. Often these messages reflect the emotion, opinion and sentiment of the public through a mix of text, image, emoticons etc. These statements are often called electronic Word of Mouth (eWOM) and are much prevalent in business and service industry to enable customers to share their point of view.

Keywords: *sentiment polarity, social media analytics, electronic word of mouth.*

GJCST-D Classification: *F.1.1*



SENTIMENTPOLARITYIDENTIFICATIONOFSOCIALMEDIACONTENTUSINGARTIFICIALNEURALNETWORKS

Strictly as per the compliance and regulations of:



Sentiment Polarity Identification of Social Media Content using Artificial Neural Networks

K. Victor Rajan^α & Brittney Jackson^σ

Abstract- Sentiment of people about consumer goods and government policies for decision making is normally collected through feedback forms, surveys etc. The social network sites and micro blogging sites are considered a very good source of information nowadays because people share and discuss their opinions about a certain topic freely. With the increased use of technology and social media, people proactively express their opinion through social media sites like Twitter, Facebook, Instagram etc. A social media sentiment analysis can help companies to understand how people feel about their products. On the other hand, extracting the sentiment from social media text is a challenging task due to the complexity of natural language processing of social media language. Often these messages reflect the emotion, opinion and sentiment of the public through a mix of text, image, emoticons etc. These statements are often called electronic Word of Mouth (eWOM) and are much prevalent in business and service industry to enable customers to share their point of view. We cannot use standard natural language processing tools to analyze the sentiment. Labeling the sentiment as positive, negative or neutral is a classification problem. We propose to use Artificial Neural Networks to identify the sentiment polarity of social media content. Without getting into the complexity of understanding the eWOM, we use machine learning algorithm to solve this classification problem.

Keywords: *sentiment polarity, social media analytics, electronic word of mouth.*

1. INTRODUCTION

Social media analytics provides useful insights about public opinion to various decision makers like commercial organizations and government authorities. People openly discuss their opinion using short messages in social media sites. A simple tally of social mentions tells you how much people are talking about a product online. But, what are they saying? Social media sentiment analysis helps you answer this question. Rather than a simple count of mentions or comments, sentiment analysis considers emotions and opinions. It involves collecting and analyzing information people share in social media. Ongoing social media sentiment analysis can quickly alert someone when customer preferences and desires change. Social sentiment analysis tools can help to ensure that you are on top of changes in what your audience expects from your brand. First, it can alert service and support teams about any new issues they should be aware of. Then,

the company can prepare a proper response, strategy or script. Organizations may even learn about issues with their product or service. Second, monitoring for social mentions with negative sentiment allows owners to reach out to people who may be having a challenging experience with their product. A simple response or follow-up can often go a long way to resolve a customer complaint. A traditional opinion poll would involve surveys with many people traveling, making phone calls etc., resulting in thousands of dollars to run. If we develop a method to identify sentiment of social media content, then this can be used to identify events of potential impact to decision makers like government agencies, commercial organizations etc. For example, online advertisers could use this analysis for efficiently targeted marketing campaigns. Government organizations can know how the society is influenced by a decision or policy and then determine how to respond to the public opinion. In addition to large volume, social media also gives the facility to collect messages pertaining to a specific region, language, a historical period etc. With the dramatic increase of social media usage nowadays, we can use the opportunity to analyze the text rich blogs and messages in a faster and cost-effective way than traditional polls. The research on social media analysis is still evolving. It might be just impossible to write natural language parsers to understand the context of these messages. We try to classify the sentiment polarity of the message as positive or negative or neutral. Hence identifying the sentiment polarity is a classification problem. Understanding the language of eWOM becomes complicated due to the non-standard structure of messages. Our objective is to develop a system which will find out sentiment polarity of people over a topic. First, we try to extract the text content after removing non-textual content like emoticons and images. Then the message is sent to a well-trained Convolutional Neural Network (CNN) for classification. The neural network classifies the sentiment as Positive or Negative or Neutral.

The remainder of this paper is organized as follows. In section 2, we discuss related work. We formally formulate the problem and we propose our methodology in Section 3. Section 4 describes how CNN is used to predict sentiment polarity and the results of our experiment are described. Finally, the paper is concluded in Section 5.

Author α: Department of Computer Engineering, Atlantic International University, Honolulu, USA. e-mail: victor@jts.co.in

II. RELATED WORK

Prediction of popular events from news and social media has been done by many people in the past. Twitter is significantly different from most text streams (e.g., news stream and blog stream) that have been extensively studied in the literature for bursty feature/event detection because of its informal writing style and topic diversity. A simple statistic in our study shows that the number of distinct bursty segments is about 75% of the number of distinct tweets in a randomly chosen time window. Among the bursty segments detected, many contain misspelled words, informal abbreviations and emojis. These noisy bursty segments would not only incur unnecessary computational cost but also hurt the event detection accuracy in further processing. We therefore source for the wisdom of the crowds to filter out the non-textual content. In this section, we present an overview of the previous works done in the field of event detection and extraction. Sentiment analysis algorithms fall into one of three buckets:

- i) *Rule-based*: These systems automatically perform sentiment analysis based on set of manually crafted rules.
- ii) *Automatic*: These systems rely on machine learning techniques to learn from data.
- iii) *Hybrid*: These systems combine both rule-based and automatic approaches.

Usually, a rule-based system uses a set of human-crafted rules to help identify subjectivity, polarity or the subject of an opinion. These rules may include various NLP techniques developed in computational linguistics such as stemming, tokenization, part-of-speech tagging and parsing.

Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral.

Hybrid systems combine the desirable elements of rule-based and automatic techniques into one system. One huge benefit of these systems is that results are often more accurate. We analyzed machine learning approaches used by researchers in the past. The text classification using ML approach can be roughly divided into supervised and unsupervised learning methods. The supervised methods make use of large number of labelled documents. The unsupervised methods are used when it is difficult to find these labelled training documents. Chi-square method is used in many applications; one of them is contextual advertising as presented by Fan and Change [3]. They discovered bloggers' immediate personal interests in order to improve advertising. They used Support Vector

Machines (SVM) for classification and Chi-square for Feature Selection (FS). Hagenau and Liebmann [4] used feedback features by employing market feedback as part of their feature selection process regarding stock market data. They used SVM as classifier. The feature selection scheme proposed by Duric and Song [5] achieved competitive results for document polarity classification specifically when using the syntactic classes and reducing overlaps with semantic words. Maximum Entropy (ME) classifier was used by Kaufmann [6] to detect parallel sentences between any two language pairs with small amount of training data. Their results showed that ME classifier can be produce useful results for almost any language pair.

SVMs are used in many applications for classification problems. Sentiment polarity identification is a classification problem. Chen and Tseng [7] have used multi class SVM based approaches. SVM have been extensively and successfully used as a sentiment learning approach while Artificial Neural Networks (ANN) have rarely been considered in comparative studies in the sentiment analysis literature. Moraes and Valiati [9] presented an empirical comparison between SVN and ANN for document level sentiment analysis. Van de Camp and Van den Bosch [10] showed that SVM and ANN can be used to mark the relationship between two persons (positive, negative or unknown). They also proved that SVM and one layer NN achieve the highest scores. The supervised learning algorithms when combined with ANN achieve good results for classification problems. However, we need sufficiently large data set with labelled classifiers for training. Though the social media sites emit huge number of messages every day, attaching labels and creating training data set is still a challenging task due to the noisy and unstructured content. Automated NLP tools also cannot be used to attach labels to them. Creating a sufficiently large training data set by manually labelling social media messages is cumbersome. So, we propose a novel approach of using semi-supervised artificial neural networks to process the social media content.

III. METHODOLOGY

In sentiment analysis, 'sentiment polarity' takes a context sensitive meaning. Rule based systems calculate the sentiment polarity based on the net of positive and negative opinion expressed about an event but fail to include the context of the event. In general, we derive day-to-day sentiment scores by counting positive and negative messages. A message with more negative words (reflecting anger, sad, violence) than positive (reflecting happy, joy, celebration) is considered as negative polarity. Application of a lexicon involves calculating the sentiment from the semantic orientation of word or phrases that occur in a sentence. With this

approach a dictionary of positive and negative words is required and a positive or negative sentiment value assigned to each of the words. Different approaches to create dictionaries have been proposed, including manual counting and automatic approaches. In lexicon-based approaches a piece of text message is represented as a bag of words. Following this representation of the message, sentiment values from the dictionary are assigned to all positive and negative words or phrases within the message. A combining function, such as sum or average, is applied in order to make the final prediction regarding the overall sentiment for the message. For example, sentence 'Enjoying my lazy Sunday!!' represents a positive message that contains one positive (enjoying) and one negative (lazy) word. It will be difficult for rule-based classifier to decide between positive and negative in such a case. A tweet message may be even worse for parsing because it might contain emoticons and special symbols. To alleviate this issue, we use machine learning to predict polarity and the decision making is similar to human behavior. Sentiment analysis of social media content is a classification problem. KNN is widely used for classification problems. But it requires sufficiently large amount of labeled data set for training. Labelling the social media content and generating few millions of training data is cumbersome due to the varying style of micro bloggers. In this paper, we propose a semi-supervised method using Neural Networks. Deep learning is very influential in both unsupervised and supervised learning.

a) Feature extraction from eWOM

The social media messages are not plain text messages. The eWOMs often contain emoticons, abbreviations etc. Sentiment analysis will not be accurate unless we translate these messages into plain text without losing the context and emotions attached. For example, the context of the sentence 'Enjoying my lazy Sunday 😊' cannot be understood unless we know the meaning of 😊. It has two positive (enjoy, 😊) and one negative (lazy) word. The local context of the message will be lost unless we take the emoticons also into consideration. So, our first step is to develop a feature extractor which will produce plain sentence from eWOM. The feature extractor works as follows.

- We pick up only English text message. Other language sentences are removed.
- Emojis are replaced with CLDR meaning, if mapping is available (Table 1). Removed otherwise.
- All words are converted to lowercase.
- Stop words and punctuation like periods, commas and brackets are removed.
- Remove all words that are not purely comprised of alphabetical characters (words having special characters and numbers).

The cleaned-up sentence is now used for identifying the sentiment polarity.

Following table shows example of few emojis translated using Unicode Common Locale Data Repository (CLDR).

Table 1: CLDR translation of emotions

| UNICODE | EMOJI | CLDR MEANING |
|---------|-------|---------------|
| U+1F600 | 😄 | Grinning face |
| U+1F642 | 😊 | Smiling face |
| U+1F609 | 😉 | Winking face |
| U+1F615 | 😕 | Confused face |
| U+1F622 | 😭 | Crying face |
| U+1F602 | 😂 | Tears of joy |
| U+270C | ✌️ | Victory hand |

For example, the tweet message 'Salute to our warriors 🤖 who taught us to raise voice against evil!' is converted into the plain text 'Salute to our warriors with tears of joy who taught us to raise voice against evil' using our feature extractor.

We created a limited data set from social media for supervised learning. Using movie review or customer review data set for social media sentiment analysis does not provide high accuracy prediction. A hand-crafted training data set consisting of 2000 positive, 2000 negative and 2000 neutral messages which were classified manually by experts was used to train the neural network. Our model learns to associate a given

input (sentence) to the corresponding output (polarity) based on the test samples used for training. Pairs of feature vectors and tags (e.g. positive, negative, or neutral) are fed into the machine learning algorithm to generate a model. The well-trained model generates predicted tags (positive, negative, or neutral) for a given input.

b) Polarity Sensitive Convolutional Neural Network (PSCNN)

Deep learning is prosperous because of three main and important reasons, i.e., improved abilities of chip processing (GPU units), extensively lower

expenditure of hardware and significant enhancements in machine learning algorithms. It is a promising approach and has been extensively applied in artificial intelligence fields like computer vision, transfer learning, semantic parsing, natural language processing and many more. People use deep neural network architecture to evaluate the similarity of documents. We present a new architecture for sentiment analysis which operates directly at the sentence level and uses only small convolutions and pooling operations. We will use a pre-trained word embedding prepared on a very large text corpus. We propose a Polarity Sensitive CNN (PSCNN) for eWOM sentiment modeling. The PSCNN is a hierarchical model, where feature extractor formats sentence vectors which are fed to the convolution and max-pooling layers to generate the document representation. Skip - gram models is best suited for context-based prediction. The model learns each term within a

given context window in the order of a word sequence to capture the skip-gram based contextual features. Then, based on the significance learned from the skip-gram features, the PSCNN structures the feature vectors up onto the sentence level. As a result, the model applies a nonlinear transformation to generate a continuous vector representation for the entire text corpus by extracting the high-level semantic information. Convolution in the proposed model is followed by global max-pooling. use sequence of word embeddings trained on large data collections as inputs to train a CNN-based representation learning model where each sequence of k words is compacted in the convolutional networks. Following diagram shows our two-step methodology which overcomes the natural language processing challenge through feature extractor and CNN.

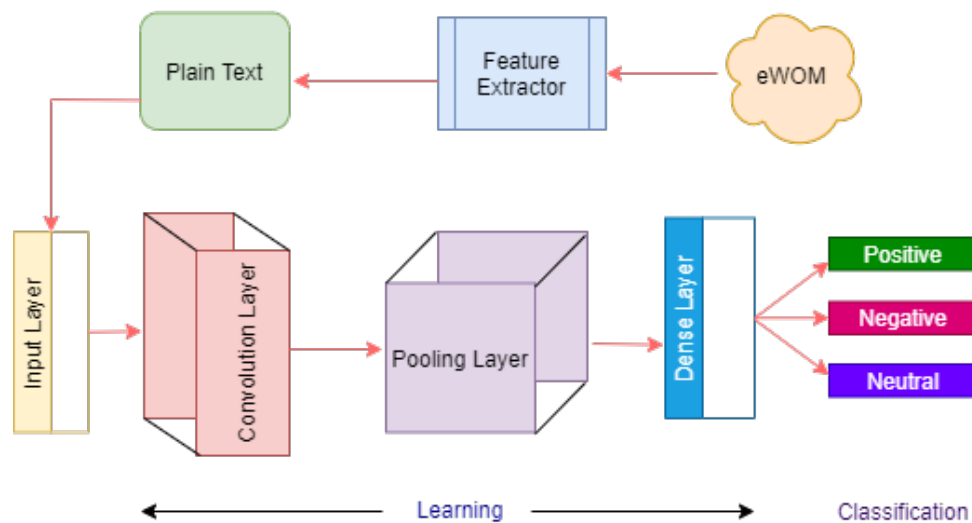


Figure 1: Hierarchical model for PSCNN

The purpose of the model was to match two sentences and to serve the paraphrasing tasks originally. PSCNN first learns and extracts representations from the two sentences separately, and then it compares the extracted features with max layer pooling to generate a matching degree. Following are the major steps involved in the convolution layer.

- The first step is to load the word embedding as a directory of words to vectors.
- The model performs a 1-D convolutional operation to learn text representations.
- This representation makes it possible for words with similar meaning to have a similar representation.

The trained model produces an output -1 (negative) or 0 (neutral) or 1 (positive).

IV. EXPERIMENTS AND RESULTS

Sentiment Analysis using Artificial Intelligence makes the life of commercial product organizations

easier. A company which sells hundreds of products cannot afford to employ many reviewers to read all the customer reviews manually and classify them. Sentiment analysis can identify critical issues in real-time, for example:

- Is a public crisis on social media escalating?
- Is an angry customer about to churn? etc.

Artificial Intelligence based Sentiment Analysis is really an efficient and cost-effective tool for many commercial organizations and government agencies. We used twitter data related to product reviews for our experiments and analysis.

In this section, we describe the evaluation tasks, the data sets used and the experimental results of the proposed approach.

Evaluation Tasks:

- We evaluate our approach on twitter data set.
- We evaluate our approach on popular benchmarks.

Our goal is to prove the real-life application of our approach. We aim to evaluate the quality of sentiment classification. The empirical probability is compared with real time data to harness the accuracy. The results show promising output.

a) Data Set

Twitter is a micro blogging site used by public nowadays to openly express their opinions and sentiment. Every day more than 500 million tweets are

generated by people around the world and this text rich social media platform serves as a desirable platform to analyze the information from many perspectives like politics, elections, consumer products and many more. Twitter API provides facility to search for messages using filters like place, language etc. Our data set for evaluation was set of tweets from May 2021 to July 2021 in English language related to India region.

Table 2: Sentiment in twitter stream

| DESCRIPTION | MAY 2021 | JUNE 2021 | JULY 2021 |
|-----------------------------------|-----------------|-----------------|-----------------|
| Total number of tweets | 5,423,667 | 5,212,765 | 5,314,443 |
| Tweets with sentiment | 1,789,810 (33%) | 1,563,830 (30%) | 1,647,477 (31%) |
| Tweets with sentiment and hashtag | 759,313 (14%) | 781,915 (15%) | 690,888 (13%) |

From the above table, we observe that appreciable number of tweets with sentiment are available for analysis. The below figure shows that tweet messages always reflect sentiment at any point in time.

We observed that the volume of positive and negative sentiments might vary over the time, but people always express their sentiment through tweets.

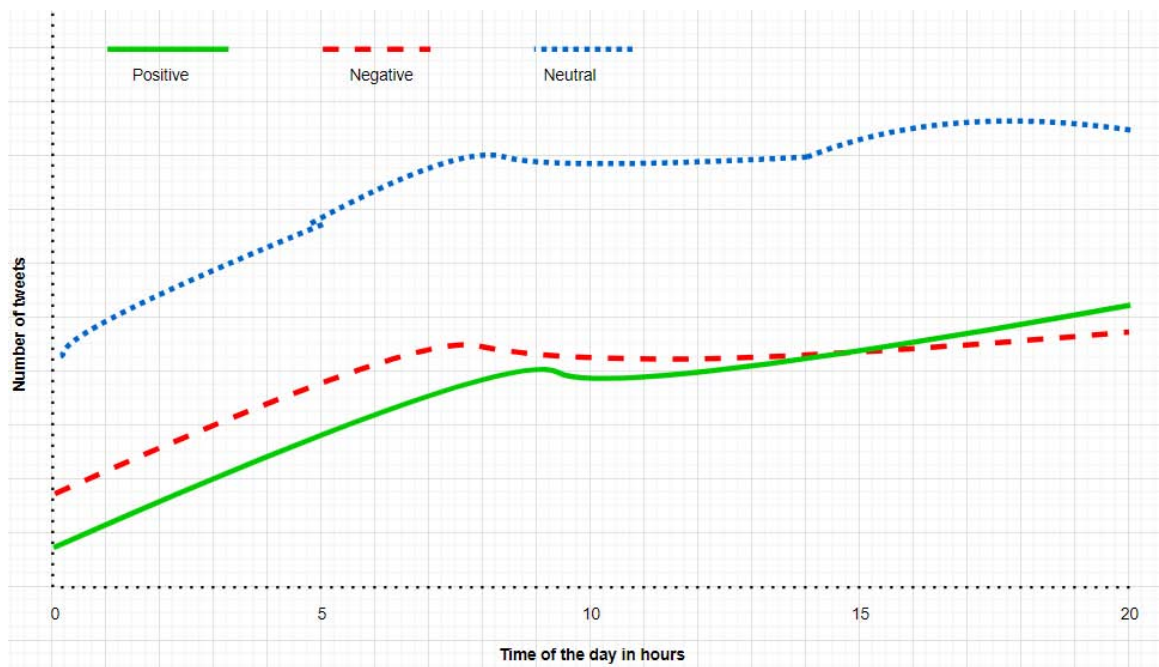


Figure 2: Distribution of sentiment in tweets

b) Popular Metrics

In this section, we first describe a set of metrics commonly used for evaluating performance of our model and then present a quantitative analysis of the performance evaluated using popular benchmarks.

Accuracy and Error Rate: These are primary metrics to evaluate the quality of a classification model. Let TP, FP, TN, FN denote true positive, false positive, true negative, and false negative respectively. The classification Accuracy and Error Rate are defined in Eq. 1

$$\text{Accuracy} = \frac{(TP + TN)}{N}, \quad \text{Error rate} = \frac{(FP + FN)}{N} \tag{1}$$

Where N is the total number of samples. Obviously, we have **Error Rate = 1 – Accuracy**.

Precision, Recall and F1 Score: These are also primary metrics and are more often used than accuracy or error rate for imbalanced test sets. Precision and recall for binary classification are defined in Eq. 2. The F1 score is

the harmonic mean of the precision and recall, as in Eq. 2. An F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad \text{F1-score} = \frac{2 * \text{Prec} * \text{Rec}}{\text{Prec} + \text{Rec}} \tag{2}$$

For multi-class classification problems, we can always compute precision and recall for each class label and analyze the individual performance on class labels

or average the values to get the overall precision and recall.

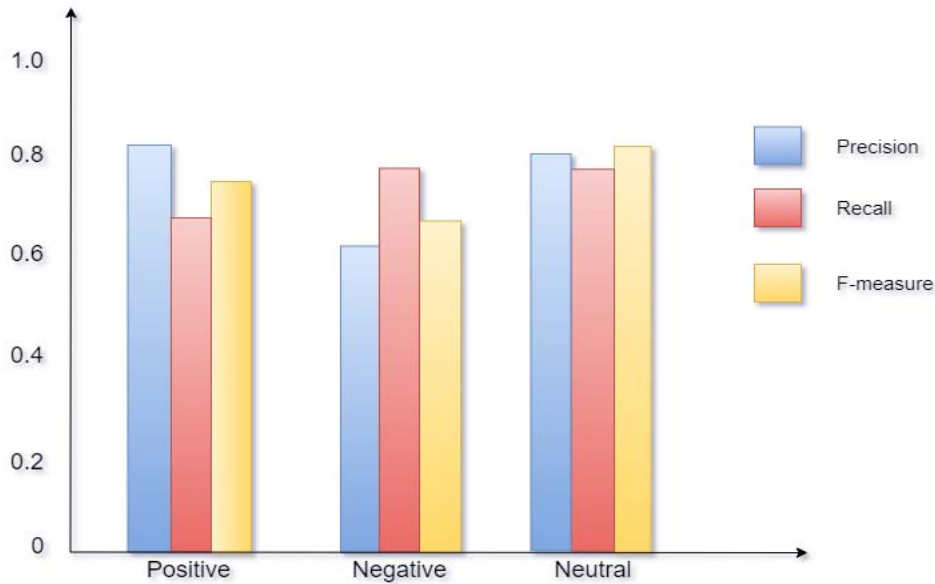


Figure 3: Quantitative metrics analysis

From the above results it can be noticed that for twitter eWOMs better accuracy is achieved at message level sentiment.

c) *Empirical Results*

Our aim is to make use of our PSCNN to analyze social media and identify impactful events

trending in the twitter. Following are examples of sentiment-full events identified using our PSCNN.

Table 3: Events identified from twitter stream

| No. | TWITTER eWOM | SENTIMENT | HASHTAG |
|-----|---|-----------|--------------------|
| 1 | Samsung Galaxy M12 is out👏. This Phone is totally Outstanding 😊! | Positive | #SamsungGalaxyM12 |
| 2 | Salute to our warriors who taught us to raise voice against evil 🙏! | Positive | #MartyrsDay |
| 3 | Their sacrifice for our motherland will continue to inspire the coming generations 🙏 We respect, value your sacrifices, we are forever indebted to your sacrifices 🙏 | Positive | #MartyrsDay |
| 4 | She is the hope of Uttar Pradesh 🙏, the only leader who is consistently raising the issues and fighting for the people everyday 😊. Young charismatic, with | Positive | #UPKiUmeedPriyanka |



| | | | |
|---|--|----------|--------------------|
| | impeccable credentials and history 🙌! | | |
| 5 | The people of UP are urging for relief from gundarj, terror & hypocrisy of Saffron reign 😞 | Negative | |
| 6 | Paranoid, vindictive government will not let farmers survive 😞. Through attacks on farmers the government has finally declared that there is emergency in India now 🙄 | Negative | #FarmersParliament |
| 7 | Farmers will discuss about APMC mandis in today's Parliament of farmers 😊. Groups of 200 farmers will protest outside the Parliament every day 🙄, during the monsoon session, to strengthen the voice in the temple of Democracy 😞 | Negative | #FarmersParliament |

From the above table, we observe that people talk about various events and express their sentiment in social media. This is really an alternate to traditional polling and cost effective solution for decision makers to understand the situation and respond to any emerging crisis.

V. CONCLUSION AND FUTURE WORK

The proposed method accomplished superior performance in terms of sentiment classification of eWOMs according to polarity. The major challenge in using the NLP tools for understanding the social media messages is eliminated by our two-step methodology namely feature extractor and CNN. By using a centralized sentiment analysis system, commercial organizations can improve accuracy and gain better insights while analyzing customer feedback and complaints. The overall benefits of AI based sentiment analysis include:

Sorting Data at Scale: Manually sorting through thousands of tweets, customer support conversations, or surveys is complex and time consuming. AI based Sentiment analysis helps businesses process huge amounts of data in an efficient and cost-effective way.

Real-Time Analysis: The social media analysis can help organizations immediately identify alarming situations and they can act right away before customer churn out.

Tagging text by sentiment is highly subjective, influenced by personal experiences, thoughts and beliefs. Intensifiers refer to words such as very, quite, most etc. These are the words that change sentiment of the neighboring non-neutral terms. They can be divided into two categories namely amplifiers (very, most) and down toners (slightly) that increase and decrease the intensity of sentiment, respectively. Identifying intensity of emotion may not be simple through rule based approached. Our AI based model can be enhanced further to identify intensity of emotion.

REFERENCES RÉFÉRENCES REFERENCIAS

- Sanjiv R. Das, Mike Y. Chen, (2007) Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science* 53(9):1375-1388.
- Walaa Medhat, Ahmed Hassan, Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* (2014) 5, 1093-1113.
- Fan Teng - Kai, Chang Chia-Hui. Blogger-centric contextual advertising. *Expert Systems with Applications* 2011; 38:1777-88.
- Michael Hagenau, Michael Liebmann, Dirk Neumann. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*; 2013.
- Duric Adnan, Song Fei. Feature selection for sentiment analysis based on content and syntax models. *Decision Support Systems*; 2012; 53:704-11.
- Kaufmann JM. JMax Align: A Maximum Entropy Parallel Sentiment Analysis Tool. In proceedings of COLING'12; Demonstration papers, Mumbai 2012. Pages 277-88.
- Chin Chen Chien, Tseng You-De. Quality evaluation of product reviews using an information quality framework. *Decision Support Systems* 2011; 50:755-68.
- Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow and Rebecca Passonneau. Sentiment Analysis of Twitter Data. Proceedings of the Workshop on Language in Social Media (LSM 2011), pages 30–38, Portland, Oregon, 23 June 2011.
- Moraes Rodrigo, Valiati Joao Francisco, Gaviao Neto Wilson P. Document level sentiment classification: an empirical comparison between SVM and ANN. *Expert Systems with Applications* 2013; 40:621-33.
- M Van de Camp, A Van den Bosch. The socialist network. *Decision Support Systems*, 53 (4), 761-769, 2012.

11. Kumar Ravi, Vadlamani Ravi. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge - Based Systems, Volume 89, 2015, Pages 14-46. ISSN 0950 - 7051
12. Nádia F.F. da Silva, Eduardo R. Hruschka, Estevam R. Hruschka. Tweet sentiment analysis with classifier ensembles. Decision Support Systems, Volume 66, 2014, Pages 170-179. ISSN 0167-9236.
13. Isa Maks, Piek Vossen. A lexicon model for deep sentiment analysis and opinion mining applications. Decision Support Systems, Volume 53, Issue 4, 2012, Pages 680-688. ISSN 0167-9236.
14. Anna Jurek, Maurice D. Mulvenna and Yaxin Bi. Improved lexicon-based sentiment analysis for social media analytics. Security Informatics – A springer open journal. (2015) 4:9.
15. A. Severyn and A. Moschitti, Twitter Sentiment Analysis with Deep Convolutional Neural Networks, Proc. 38th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR 15, pp. 959-962, 2015.
16. S. Zhou, Q. Chen, and X. Wang, Active deep learning method for semi-supervised sentiment classification, Neurocomputing, vol. 120, pp. 5365-46, 2013.





Human Tracking and Profiling for Risk Management

By Ranjith K.H.V.S, Jayasekara A.S, Ratnasooriya K.A.L.L., J.L Thilini Randika,
Dr. Lakmal Rupasinghe & Ms. Chethana Liyanapathirana

Institute of Information Technology

Abstract- Infectious viruses are conveyed via respiratory droplets produced by an infected person when they speak, sneeze, or cough. So, to combat virus transmission, the World Health Organization (WHO) has imposed severe regulations such as mandatory face mask use and social segregation in public spaces. The 'Human Tracking and Profiling for Risk Management System (HTPRM)' is an online application that identifies the risk associated with failing to follow proper health practices. This proposed approach, which is divided into four components, utilizes 'You Only Live Once YOLO (V3)' to detect face-mask danger, which would be determined based on two factors: wearing the face mask properly and the type of mask (Surgical, k95, homemade, and bare). The second phase is to use Open CV and SSD-Mobilenet to evaluate the value of a one-meter space (Social Distance) between people. The system recognizes the maximum number of individuals that can be in the vicinity of the specific hall that uses YOLO(V3) and image processing as the third procedure. In the last processing, the system identifies each person's behavior, classifies it as uncommon or not, and calculates the risk associated with each category. Finally, the system computes the overall risk and generates a warning alarm to notify the user that they are in a dangerous scenario.

Index Terms: YOLO (V3), SSD (single shot detector), mobile-net, open-CV, image processing, open pose, tensor-flow.

GJCST-D Classification: H.3.4



HUMAN TRACKING AND PROFILING FOR RISK MANAGEMENT

Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

Human Tracking and Profiling for Risk Management

Ranjith K.H.V.S ^α, Jayasekara A.S ^σ, Ratnasooriya K.A.L.L. ^ρ, J.L Thilini Randika ^ω, Dr. Lakmal Rupasinghe[¥] & Ms. Chethana Liyanapathirana[§]

Abstract Infectious viruses are conveyed via respiratory droplets produced by an infected person when they speak, sneeze, or cough. So, to combat virus transmission, the World Health Organization (WHO) has imposed severe regulations such as mandatory face mask use and social segregation in public spaces. The 'Human Tracking and Profiling for Risk Management System (HTPRM)' is an online application that identifies the risk associated with failing to follow proper health practices. This proposed approach, which is divided into four components, utilizes 'You Only Live Once YOLO (V3)' to detect face-mask danger, which would be determined based on two factors: wearing the face mask properly and the type of mask (Surgical, k95, homemade, and bare). The second phase is to use Open CV and SSD-MobileNet to evaluate the value of a one-meter space (Social Distance) between people. The system recognizes the maximum number of individuals that can be in the vicinity of the specific hall that uses YOLO (V3) and image processing as the third procedure. In the last processing, the system identifies each person's behavior, classifies it as uncommon or not, and calculates the risk associated with each category. Finally, the system computes the overall risk and generates a warning alarm to notify the user that they are in a dangerous scenario.

Index Terms: YOLO (V3), SSD (single shot detector), mobile-net, open-CV, image processing, open pose, tensor-flow.

I. INTRODUCTION

Viruses have been a point of controversy for humans since even before our species evolved into its contemporary form. Vaccines and antiviral drugs have allowed to limit the widespread spread of certain viral diseases and have helped sick patients to recover. Some viruses, such as Marburg, Ebola, Rabies, HIV, Smallpox, SARS - Cov, and MERS - Cov, have been unable to be eradicated, leading to a rise in new cases. However, destroying viruses is a challenging task. Some viruses have migrated from animals in recent decades, triggering large epidemics and claiming thousands of human lives. From 2014 to 2016, the virus that caused the Ebola epidemic in West Africa killed approximately 90% of those infected, making this the most dangerous member of the Ebola family. But other viruses are just as deadly, and some are even more

deadly. It is true that the death rate from some viruses, such as the new corona-virus, which is currently spreading around the world, is still high, but infections pose a serious threat to public health as human cannot yet combat germs. The goal of the HTPRM is to identify risk by investigating at four different virus transmission modes. Crowd Counting is one technique that determines the population of a scene. Public places are overcrowded possible to spreading the virus among peoples. To minimize the spreading viruses, must be limited the number of people in a public area. The density risk percentage calculate by comparing the number of people in a specific area at a given moment to the maximum number of people permitted in that region. This is one method to help prevent virus propagation in densely populated places, since it is a highly robust strategy in today's society.

The primary source of risk is that COVID-19 spreads by touch or being in close proximity to an infected individual. Social Distancing is the only way to prevent the spread of COVID-19. Maintaining a safe distance from one another is the most effective method of preventing the transmission of this disease at least until a vaccine is found. Social distancing can have a negative impact on social well-being and health by resulting in social isolation and physical activity limitations.

One of the another critical of risk is behavior control. If one person is infected with a virus, there is a high possibility of the infection spreading due to his behavior. To avoid or mitigate such a circumstance, a human behavior recognition function has been developed for risk management systems using Human Tracking and Profiling. This capability is mostly concerned with human behavior in a given area.

Another critical aspect of combating this epidemic is to use a face mask. Wearing a face mask has received general support as a means of delaying the transmission of viruses. Speed of virus transmission is dependent on the sort of face-mask used (Surgical, k95, homemade, and bare). Until the virus is completely destroyed, daily use of face masks is crucial for infection prevention and protection against airborne infectious germs. This approach not only determines whether or not someone is wearing a mask, but also determines whether or not a face mask is worn properly. The danger varies according to the type of face mask used, the

Author ^α ^σ ^ρ ^ω [¥] [§]: Dept. of Computer Software Engineering Sri Lanka Institute of Information Technology. Malabe, Sri Lanka.
e-mails: it18126020@my.sliit.lk, it18121834@my.sliit.lk,
it18115444@my.sliit.lk, it18044928@my.sliit.lk, lakmal.r@sliit.lk,
chethana.l@sliit.lk

average risk is determined as well. This research will conduct an in-depth examination of the usage of masks to prevent the transmission of the lethal coronavirus. This version introduces a novel multi - facemask detection in real-time. Finally, the Human Tracking and Profiling system calculates the average risk associated with four components and then calculates the total average risks involved with those four components.

II. LITERATURE REVIEW

The main intention is to implement automated Human tracking and profiling for risk management application to avoid speed of spreading the virus infections in the world. Social Distance Risk, Face-mask risk, Density Risk and Human Actions and behaviors Risk are the main four components to minimize the percentage of deaths due to viruses. Lot of applications developed for analysis the risk and get the crowd/visitors count in the frame but there is not single system to analysis the density risk in a particular area to minimize the spreading viruses among people. In the current pandemic situation in world, one Deep-CNN Crowd Counting Model for enforcing social distancing application is implemented in Saudi Arabia's public places for avoid spreading the viruses among peoples [1]. Actually, above proposed method is based on CNN model to count people who appear in video frames in public places [1]. Another, people counting system developed in post COVID-19, which is counting people through infrared detection and this system count and update based on people moving in/out through the area/premise [2]. There already exists a few work that pedestrian counting systems [[3], [4]]. so, in this proposed density risk solution go beyond above systems and in the first step, a video first frame user [system owner] must selects the area where he/she wants to measure the density risk using four mouse click points. Then according to area width and length system estimate the maximum people/visitors count allowed in that area. After that, system take the real time total people/visitors count inside the user selected area (where user wants to measure the density risk) in each video frames and system comparing the these both real time people count, and maximum people count allowed in this area and analysis get density risk. If the real people count is higher than the maximum people count allowed in the area, then that area is a high-risk place. This is the novelty of density risk analysis. When it comes to the face-mask risk, the majority of the research studies reviewed focused exclusively on identifying the face mask. Researchers used a variety of machine learning and deep learning algorithms to assess whether or not they were wearing a face mask. Using image processing, the device developed by a team led by S. Balaji detected the passengers' facemasks [9]. Additionally, the team, which includes

Amit Chavda, has presented a method that uses a Convolutional Neural Network to detect individuals who use facemasks [7] and some of research papers used to detect facemask by utilizing Faster-RCNN [[6], [8], [10]]. Numerous devices have been proposed and implemented to detect facemasks using various methodologies, however analysers all have significant limitations. Numerous facemask types have been introduced to the market. Even if individuals are masked, it is impossible to demonstrate that they are passing hatred from one individual to another. This is because they must cover their nose and mouth and secure it beneath their chin, even if they are wearing a mask. Additionally, it clings snugly to their chin. The National Center for Immunization and Respiratory Diseases (NCIRD) has confirmed that the viral transmission rate varies between different types of masks. Certain masks are designed and tested to ensure consistent performance in preventing the transmission of COVID-19. These masks are labeled with the criteria they comply with. KN95 masks provide approximately 98.5% protection, whereas surgical masks provide 56.1 percent protection. Some folks make their homemade masks. This results in a 51.4% guard. We concentrated on that and identified the facemask as surgical, KN95, and homemade using the YOLO principle (V3). Thus, our proposed system analyzes the multi-person real-time face mask type and analyzes the risk of face masks and unmasking. According to the investigation, if it exceeds 75%, it is considered a risky zone. As soon as it becomes a risky area, the head of the location is warned through SMS. Another important aspect of human profiling and estimating for risk management systems is the estimation of human actions and behaviors. In the study publications, systems to detect an individual's activity were introduced, but systems to recognize the action of a group of individuals were not found. Zhe Cao and colleagues focused on a critical component of acquiring a deep understanding of humans in photos and videos: human two-dimensional posture estimation—or the difficulty of localizing anatomical important points or "parts." Human estimating has always been primarily concerned with locating individuals' body components [14]. Federico Angelini and his colleagues proposed Action Pose: a two-dimensional pose-based technique for human action recognition at the pose level [15]. They retrieved low-level and high-level features for the Action XPose from the human body posture and fed them into a LSTM (Long Short-Term Memory Neural) Network and a 1D Convolutional Neural Network for classification. Action XPose, a 2D pose-based algorithm for posture-level Recognizing Human Action, was introduced by Zeyu Fu and his team [16]. However, our suggested system recognizes the action of human and classifies actions such as leap, run, and walk. The risk is assessed based on the classification of human behavior

.To accomplish this, we use a computation algorithm that we created ourselves. As with other systems, if someone behave incorrectly, it sends a notification to the location chief.

When analyzing the risk of social distance to get a better accuracy this paper used the SSD-mobilnet model. In each frame need to be more accurate. Therefore to get better detection for bad visibility areas used CLAHE preprocessing method to identify objects [11]. Most of the researchers only considered about detect social distance [13]. This paper is most related to analyze the risk of percentage. Based on these percentages, can easily detect whether the area is bad or good.

III. METHODOLOGY

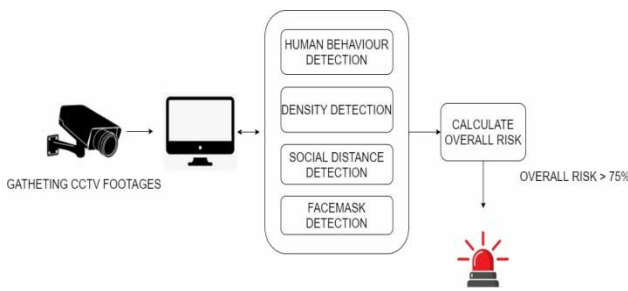


Fig. 1: Overall System diagram

According to the system overview diagram Fig.1, initially system gets CCTV footage as a input and same CCTV footage goes through four sub risk analysis functions separately and estimate the risk status. For the estimating the overall risk percentage, divide the 100% equally among four sub functions and each function gets 25%. if one function is totally violating, then added each sub function 25% percentage to the total overall risk. finally if the total risk percentage is greater than 75% email will be send, informing that area will be a risky place.

a) Density Risk Analysis

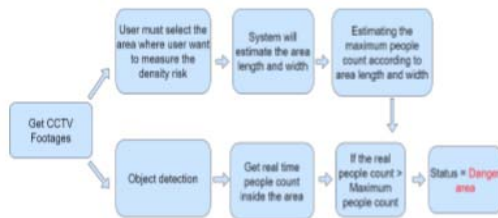


Fig. 2: An image of a density risk system overview

The proposed system analysis the density risk in a particular area at a particular time. According to Fig.2 the system gets video frames as input and then in the first video frame user must select the area where they want to measure the density risk using four mouse click points. Then this area is a Polygon shape rectangle area. Then according to the width and length of the

area, the system estimates the maximum people count allowed in that particular area according to Fig.3 using predefined formula. After that, the system takes the real-time people/visitors count inside the area in each video frame. finally, the system comparing both the Maximum people count and real-time people count inside the area, and if the real-time people count higher than the Maximum people count that area is a High-Risk area, email will be send, informing that area will be a risky place.

Object detection and tracking are one of the main parts of this function.Yolov3 is used to detect an object in the frame. Yolov3 is a unique neural network that predicts bounding boxes and class probabilities directly from complete images in a single evaluation. The Yolov3 configure (cfg file) and weight file trained on the detect 80 classes objects. However, People/visitors type object detection is only needed for the density risk estimation. So then did some transfer learning (hyper-parameter changing – max batches, filters, classes) for the Yolov3.cfg file (model architecture file) and re-trained using Google’s Open Images, then generated the new weight file and it used for the people/visitors object detection in the video frames and also shapely python libraries and Open CV techniques used for the estimate the length and width of the area in the video frame.

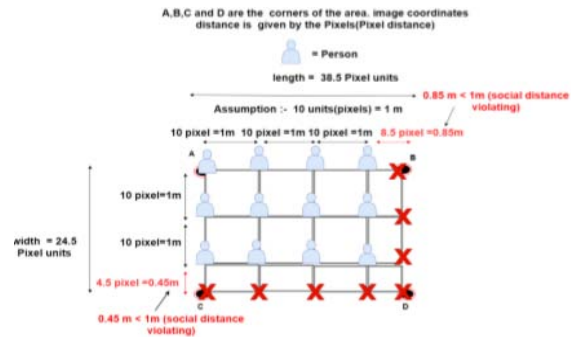


Fig. 3: Example for maximum people estimation

b) Social Distance Risk Analysis

Proposed method was developed to detect the safety distance between people in public areas. The CNN based methods such as YOLO (You look only once), SSD (Single shot Detector) computer vision and machine learning techniques are employed in this project. SSD Object Detection extracts feature map using a base deep learning network, which are CNN based classifiers, and applies convolution filters to finally detect objects. Here are the steps. Mainly the open source open-cv is used to divide video into small frames. The SSD-MobileNet-caffe model which is used to detect the objects and analyse the bounding boxes.

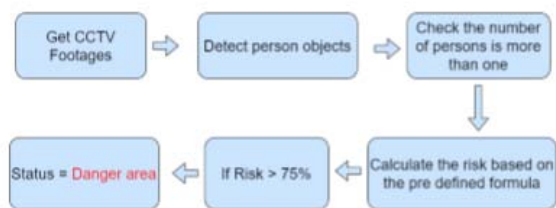


Fig. 4: Steps for detect social distance

For camera setup, It is shot from a fixed angle as a video frame. Further, the video frame was viewed as a forward view and converted to a two-dimensional view for a more accurate assessment of the distance measurement. This is the main workflow of the model we use to determine social distance.

1. Consider an image/video
2. Divided into small framed
3. Pass an image to SSD-MobileNet model
4. Extract the features of an image
5. Identify the objects
6. Used euclidean matrix to get each distance of identified bounding boxes.

This algorithm calculates distances between people and draw different colors of bounding boxes with fulfilling above steps. Used SSD-MobileNet model for object detection. For better detection for bad visibility areas used CLAHE preprocessing method.

c) *Human Behaviour Risk Analysis*

This proposed Human Tracking and Profiling for Risk Management another main part is Human behaviour recognition part. Briefly in this part, Estimate the human actions and then recognize what are the actions using previously estimated actions. In this scenario, mainly there are two main parts in the human action recognition.

i. *Human action estimation*

In this part there are also two parts, Which is, single person post estimation and multi person action estimation. Multi-person action estimation is more difficult than single person action estimation, Because, there are more than one object should be locked in the each frame. According to this proposed system we had to use multi person human action estimation method. There are lot of multi human action estimation method. for a example, Open Pose, Alpha Pose , Deep Cut and Mask RCNN. from them, Open Pose, and Deep sort algorithm methods are used to develop this function, Because, it gives more accuracy than other methods, and there are more capable facility to get real time human actions. And another advantage is, Open Pose follows Bottom-Up approach. In the bottom-Up approach, first initially detect the human joints and the

connect each joint for each related person. Deep Sort algorithm is mainly used for track multi people.

ii. *Recognize human Behaviours*

This is the second part of this function. In this function recognize the what are the human behaviours using previously estimated human actions. To do that, we used a machine learning model that we created using more than 4100 image data as the data set.

d) *Face Mask Risk Analysis*

Facemask risk was monitored in real time using a deep learning approach for detecting face masks. This section identifies the type of facemask and calculates the risk by comparing it to the recommended risk values. Two distinct YOLO (V3) object detection models are used to determine whether a face mask is present or absent and to classify the type of face mask. YOLO is an ingenious convolutional neural network (CNN) for real-time object detection. The algorithm applies a single neural network to the entire image, then divides it into regions and predicts their bounding boxes and probabilities. Here the feature maps are obtained by 81,79 and 91 convolutional neural network layers in three detections. In this detection is accomplished by applying detection kernels to feature maps of three distinct sizes located in three distinct locations throughout the network.

Due to the difficulty of obtaining a sufficiently large dataset for training the two models, custom datasets were used. A dataset of 6000 images was used to classify four types of facemasks: Surgical, KN95, Homemade, and Bare. Additionally, to determine whether or not to use facemasks, we used a dataset of 4000 images. Where transfer learning was used to train YOLO (V3) models using custom data.

Then, using Python, author created an algorithm based on the risk value assigned by the (NCIRD), Division of Viral Diseases. If the area is dangerous, the head of location is notified via SMS. Twilio's Python library assists in creating a new instance of the Message resource by allowing you to specify the message's To, From, and Body parameters.

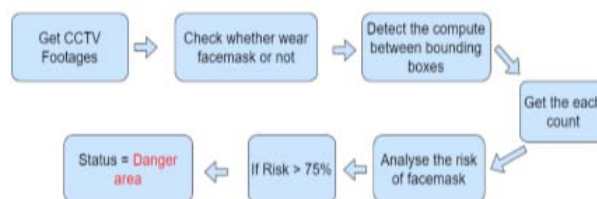


Fig. 5: Method of face-mask detection

IV. RESULTS AND DISCUSSION

a) *Density Risk Analysis*

Yolov3 network was first trained with a single class dataset of 4000 people images. So the dataset

only has one class, the maximum batch was set to 8000, the steps were set to 6400, 7200, and 18 filters in the three convolutional layers before the YOLO layers, and the number of classes in the YOLO layers was set to 1 and also set network size width 608 height 608 in Yolov3.cfg file. Map value test on 500 people images and got 85% map value for our yolov3 trained model [5]. Table I [5] compares our Yolov3 approach to a variety of different object detection methods in terms of mAP. Fig.6 shows the results of the density risk estimation. if the real-time visitors count inside the area is higher than the maximum count that area is a high-risk area and an email is sent to the nearest police station.

Table 1: Map comparisons

| Model | Dataset | MAP (%) |
|------------------|--------------------|---------|
| Our YOLOv3 | Google Open Images | 85.0% |
| Alexey AB YOLOv3 | Pascal Voc | 87.0% |
| R-CNN | Pascal Voc | 53.2% |



Fig. 6: Final output of Density Risk detection

b) Social Distance Risk Analysis

Social Distance Algorithm is a method for controlling epidemic diseases. People use social distance to protect in any epidemic circumstance. This system calculates distances between people and draws various border colors for three risk degrees. Used SSD-Mobile Net model for object detection. The accuracy of developed model SSD-Mobilenet was 92.8 percentage. Authors tested the proposed model using a video stream and images. In each frames were also labelled as unsafe and safe accordingly. To bad visibility areas proposed using CLAHE preprocessing technique. It is vital to have individuals moving continuously while utilizing the webcam, or else the detection will be wrong.



Fig. 7: Final output of social distance detection

c) Human Behaviour Risk Analysis

Another main part in this system is Human behaviour risk analysing. To estimate the human actions, we mainly used two pre-trained Open Pose

models to estimate the human actions. The main part of this function is recognizing human actions using estimated human actions. To do that we used over 4000 image data to train a model. After the training, we were able to get a 98.3 percentage training accuracy and 95.9 percentage of test accuracy.

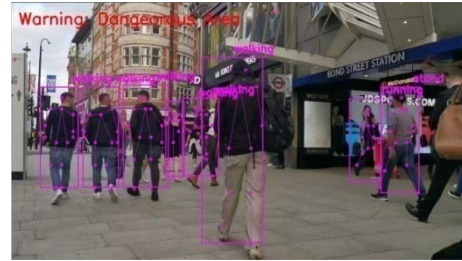


Fig. 8: Final output of Human behaviour Recognition

According to above Fig.8, we can get a clear and good idea about the result of this Human Behaviour recognizing function.

d) Face Mask Risk Analysis

The wearing of masks correctly and consistently is a vital step that everyone can take to avoid contracting and spreading COVID-19. Masks are most effective when everyone wears them, but not all masks offer the same level of protection. Consider how well a mask fits, how well it filters the air, and how many layers it has when purchasing one. For the purpose of this research, the data sets which have two classes (MASK and No Mask) and four classes (Surgical, KN95, homemade and bare) were obtained. For the facemask risk detection using facemask type, a YOLO(v3) model was pre-trained with Pytorch Geometric using custom dataset imported from YOLO v3 achieving a train mean average precision of 99.24% and test mean average precision of 73% with 6000 images in training and 2000 test images under 4 classes in validating the model. Figure 10 is shown it efficiently. With these findings, our model has also demonstrated success in detecting face masks in images beyond the our training and validation range. We initialized our learning rate at (LR=0.001), the number of training epochs at (EPOCHES = 45000), and the batch size at (BS = 64) for the testing phrase. Figure 9 depicts various scenarios for detecting different sorts of face masks in real time from a live-stream. Additionally, Table 1 discusses the importance of performance indicators in gaining a better knowledge of how suggested models behave throughout the testing process. The result analysis demonstrates that our suggested approach for face mask detection based on several types of masks performs really well despite the fact that testing data is limited.

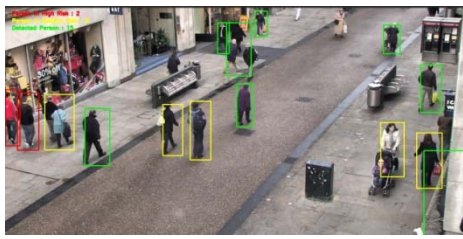


Fig. 9: Final output of Facemask detection

| Dataset | Status of Image | Mean Average Precision |
|------------------|-----------------|------------------------|
| Custom dataset 1 | Mask | 88.11% |
| | No mask | 95.07% |
| Custom dataset 2 | N95 | 99.88% |
| | Surgical | 99.90% |
| | Bare | 98.29% |
| | Homemade | 98.88% |

Fig. 10: Results of several performance key metrics based on the prediction of various training data-set

e) Overall Risk Calculation

The authors were unable to obtain individual outputs for this system after considering the above situations (Social distance Risk, Social Density Risk, Human Behavior Risk, and face mask Risk). Because different health guidelines infractions can occur in the same public space As a result, overall risk must be estimated utilizing social distance risk, social density risk, human behavior risk, and face mask risk. The authors provide a new formula to calculate the overall risk using each functionality.

$$\text{Overall Risk} = \frac{\text{Total \# of violated Risk Categories}}{\text{Total \# of Risk Categories}} * 100\%$$

Fig. 11: Overall Risk Calculation Formula

Total risk categories must be determined in order to compute the overall risk. There are four risk categories, according to the system (Social distance Risk, Social Density Risk, Human Behavior Risk and face mask Risk). The total number of risk categories that have been breached should next be determined. Finally, using these variables, compute the overall risk. If the aggregate danger exceeds 75%, the area is considered high risk. If the entire risk is between 25% and 75%, the area is considered low risk. Finally, if the threat is less than 25%, the location is considered safe.

V. CONCLUSION

The use of machine learning becomes more common. By using the image processing and deep-learning techniques, i.e. YOLO, SSD, Open Pose, and Deep-Sort methods, we provide a comprehensive real-time person recognition system. Mainly covered four main scenarios. Those are density detection and analyze the risk, social distance detection and analysis of the risk, face-mask detection and analyze the risk, and human pose detection and analyze the risk. Test average precision (mAP) for detect humans and detect

facemask with facemask type respectively 85.0 %, 73.0 %. To detect human behavior the system got 95.0% present of test accuracy. Our social distancing risk detection and estimating area length and width for density risk detection did not use correct camera calibration, which means that pixel distances to measurable real units were not (easily) mapped to (i.e., meters, feet, etc.). Therefore, the first step to improving our social distancing risk detection and estimating area length and width for density risk from the distance between our social systems is therefore to use a good camera calibration. That way, the results will be better and can calculate measurable units actually (rather than pixels). This work can be used as the basis for estimating the risk of each function. In the end, we are come up with the four individual average risks. Based on that we are calculating the total risk for a particular place.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Salma Kammoun Jarraya, Maha Hamdan Alotibi, and Manar Salamah Ali1,"A Deep-CNN Crowd Counting Model for Enforcing Social Distancing during COVID19 Pandemic: Application to Saudi Arabia's Public Places", 2020.
2. A.S.F Rahman1, S.B Yaakob1, A.R.A Razak1 and R.A Ramlee2,"Post COVID-19 implementation of a bidirectional counter with reduced complexity for people counting application", 2021.
3. Melbourne—Pedestrian Counting System. Available online:<https://www.melbourne.vic.gov.au/about-melbourne/research-and-statistics/city-population/Pages/pedestrian-counting-system.aspx> (accessed on 3 March 2021).
4. Fowzia Akhter, Sam Khadivizand, Hasin Reza Siddiquei, Md Eshrat E. Alahi and Subhas Mukhopadhyay, "IoT Enabled Intelligent Sensor Node for Smart City: Pedestrian Counting and Ambient Monitoring", 2019.
5. Nurul Iman Hassan, Fadhlan Hafizhelmi Kamaru Zaman, Nooritawati Md. Tahir, Habibah Hashim", "People Detection System Using YOLOv3 Algorithm", 2020.
6. Huaizu Jiang, Erik Learned-Miller,"Face Detection with the Faster R-CNN", IEEE 12th International Conference, 2017.
7. Zhang Jialin1, Liu Yaofeng1, Jiang Zhendong1 and Zhang Heng1,"Research on medical mask detection method based on fast Fourier transform and linear Gauss", IEEE/ASME International Conference on Advanced Intelligent Mechatronics Hong Kong, China, July 8-12, 2019.
8. Susanto, Febri Alwan Putra, Riska Analia, Ika Karlina Laila Nur, "The Face Mask Detection For Preventing the Spread of COVID-19 at Politeknik Negeri Batam", IEEE International Conference, 2021.

9. S.Balaji¹, B.Balamurugan², T.Ananth Kumar^{3*}, R.Rajmohan⁴ P.Praveen kumar, "A brief Survey on AI Based Face Mask Detection System for Public Places", A brief Survey on AI Based Face Mask Detection System for Public Places, 2021.
10. Nisarg Pethani, Harshal Vora, "FACE MASK DETECTION USING YOLO V3 ARCHITECTURE", 2020.
11. Garima Yadav, Saurabh Maheshwari, Anjali Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system", 2020.
12. [Nina B. Masters¹, Shu-Fang Shih², Allen Bukoff³, Kaitlyn B. Akel¹, Lindsay C. Kobayashi¹, Alison L. Miller⁴, Harapan Harapan^{5,6,7}, Yihan Lu⁸, Abram L. Wagne, "Social distancing in response to the novel coronavirus (COVID-19) in the United States"
13. Rinkal Keniya, Ninad Mehendale, "Real-time social distancing detector using Social distancing Net19 deep learning network"
14. GZhe Cao, Student Member, IEEE, Gines Hidalgo, Student Member, IEEE, Tomas Simon, Shih-En Wei, and Yaser Sheikh "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", 2020.
15. Federico Angelini, Student Member, IEEE, Zeyu Fu, Student Member, IEEE, Yang Long, Senior Member, IEEE, Ling Shao, Senior Member, IEEE, and Syed Mohsen Naqvi, Senior Member, IEEE "ActionXPose: A Novel 2D Multi-view Pose-based Algorithm for Real-time Human Action Recognition".
16. Federico Angelini, Student Member, IEEE, Zeyu Fu, Member, IEEE, Yang Long, Senior Member, IEEE, Ling Shao, Senior Member, IEEE, and Syed Mohsen Naqvi, Senior Member, IEEE "2D Pose-based Real-time Human Action Recognition with Occlusion-handling".



This page is intentionally left blank



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: D
NEURAL & ARTIFICIAL INTELLIGENCE

Volume 22 Issue 1 Version 1.0 Year 2022

Type: Double Blind Peer Reviewed International Research Journal

Publisher: Global Journals

Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Deep CNN Model for Non-Screen Content and Screen Content Image Quality Assessment

By Dilip Chaudhary & Venkatesh

University Visvesvaraya College of Engineering

Abstract- In the current world, user experience in various platforms matters a lot for different organizations. But providing a better experience can be challenging if the multimedia content on online platforms is having different kinds of distortions which impact the overall experience of the user. There can be various reasons behind distortions such as compression or minimal lighting condition while taking photos. In this work, a deep CNN-based Non-Screen Content and Screen Content NR-IQA framework is proposed which solves this issue in a more effective way. The framework is known as DNSSCIQ. Two different architectures are proposed based upon the input image type whether the input is a screen content or non-screen content image. This work attempts to solve this by evaluating the quality of such images.

Index Terms: *deep learning, convolutional neural network (CNN), screen content image (SCI), image quality assessment (IQA), no-reference IQA (NR-IQA).*

GJCST-D Classification: *F.1.1*



Strictly as per the compliance and regulations of:



© 2022. Dilip Chaudhary & Venkatesh. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BYNCND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Deep CNN Model for Non-Screen Content and Screen Content Image Quality Assessment

Dilip Chaudhary^α & Venkatesh^σ

Abstract In the current world, user experience in various platforms matters a lot for different organizations. But providing a better experience can be challenging if the multimedia content on online platforms is having different kinds of distortions which impact the overall experience of the user. There can be various reasons behind distortions such as compression or minimal lighting condition while taking photos. In this work, a deep CNN-based Non-Screen Content and Screen Content NR-IQA framework is proposed which solves this issue in a more effective way. The framework is known as DNSSCIQ. Two different architectures are proposed based upon the input image type whether the input is a screen content or non-screen content image. This work attempts to solve this by evaluating the quality of such images. The proposed model can be used for both screen content as well as non-screen content image to calculate the image quality score. The performance of the proposed framework was calculated using different datasets like LIVE IQA, LIVE MUL and IVL and it is found that the model is giving state-of-the-art results for all these datasets. Experiments were conducted to find the performance results for various distortion types in LIVE IQA dataset.

Index Terms: deep learning, convolutional neural network (CNN), screen content image (SCI), image quality assessment (IQA), no-reference IQA (NR-IQA).

I. INTRODUCTION

Image quality assessment is a subject of extensive analysis over the last four decades. Different multimedia applications streaming images and videos like Netflix, Amazon Prime Video, Twitter, Face book, Share Chat, etc. are gaining more popularity day by day. With the increasing availability of Internet all over the world, the usage of these applications is increasing rapidly. So, these applications requires quality assessment to be done on their content so that they can provide quality content on their platform. This helps to improve customer visual experience on their respective plat- forms. The main aim of image quality assessment is to quantitatively measure the perceived quality of digital and natural photographs. The acquisition, transmission, storage, post-processing, or compression of images brings different distortions, such as Gaussian blur (GB), Gaussian white noise (WN), or blocking artifacts. WN is added while taking pictures at night with a mobile, GB occurs if not focusing correctly before taking the shot.

*Author α σ: University Visvesvaraya College of Engineering.
e-mails: dilipcd1997@gmail.com, venkateshm.uvce@bub.ernet.in*

Based on IQA results, decisions can be taken on compression ratio for these digital images before storing them in servers for streaming purpose as well as deciding which image will be good to be published on the online platform. A dependable IQA technique can help assess the quality of photos downloaded from the web, as well as measure the accuracy of image processing techniques precisely, such as super-resolution and image compression from a human's perspective. The IQA algorithms are categorized into 3 groups, based upon the usage of reference image: no reference IQA (NR-IQA), reduced-reference IQA (RR-IQA) and full-reference IQA (FR-IQA). The performance of these algorithms is NR-IQA, RR-IQA, and FR-IQA, in order of increasing accuracy. However, since pristine images are not available in most of the real time situation, NR-IQA is most suitable method. The image quality assessed using no-reference (NR) IQA algorithms does not require knowledge of the original image. The image quality assessed using reduced-reference (RR) IQA methods requires only a few details about the original image. Full-reference (FR) algorithms need both a distorted image and a reference image as input and produce a quality rating for the distorted image in comparison to the original image. The most common technique to FR-IQA is to first calculate the local pixel-wise differences between reference image and distorted image. Finally, combine these local calculations into a single scalar value to represent the overall quality difference. Example of FR-IQA algorithms are: Structural Similarity Index Mean (SSIM), the peak signal-to-noise ratio (PSNR) and mean-squared error (MSE). Unlike FR-IQA, in NR-IQA the quality is measured using the features obtained from the distorted images and the subjective quality scores.

II. RELATED WORK

This section provides a brief detail of the existing no-reference and reference image quality assessment techniques. Li et al. [1] proposed a new multiscale directional transform, basically a shearlet transform used to extract simple features from distorted images. Then these primary features are used to explain the nature of original images and distorted images.

Then, stacked autoencoders are used to amplify the primary features and make them more distinguishable.

Mittal et al. [2] proposed a NSS-based distortion-generic IQA model. This model works best in the spatial domain. BRISQUE does not calculate the distortion-specific features, such as blur, blocking, or ringing. Rather, it uses scene statistics of locally normalized luminance coefficients to quantify losses of naturalness in the image.

Li et al. [3] trained a general regression neural network (GRNN) to assess the quality of image, relative to the human subjective opinion, across a diverse range of distortion types. The features used for assessing the quality of the image include gradient of the distorted image, entropy of phase congruency image, mean value of the phase congruency image, and entropy of the distorted image.

Moorthy and Bovik [4] introduced DIIVINE (Distortion Identification-based Image Verity and Integrity Evaluation). This algorithm evaluates the quality of a distorted image without the original images. It is a 2-stage based technique where image distortion identification is done first and then image quality assessment is done based on distortion type.

Tang et al. [5] presented a framework, where potentially neither the degradation process nor the ground truth image is known. The method is based on a set of low-level image features. The image quality characteristics are derived from original image measurement and texture statistics. Here, a machine learning technique is used to learn a mapping from these features to the subjective quality scores.

Doermann et al. [6] obtained the basic feature set by the extraction of local features. Then, using the features from the CSIQ database, by adopting K-means clustering, the codebooks with 100 centers was retained. In the mean time, the method proposes high order features: variance, mean, and skewness. The input features are used to get distances to K clusters. Then the method performs regression over three distances. It is sensitive to diverse distortion types.

Fang et al. [7] proposed a quality assessment methodology based on statistical structural and luminance features (NRSL). The evaluations were done on 4 synthetically and 3 naturally distorted image datasets. In terms of high correlation with human subjective judgments, the employed NRSL metric compares favorably to relevant BIQA models. Support vector regression was used to establish the complex nonlinear relationship between feature space and quality score. It was unable to use NRSL for various distortions in chromatic component of the image.

Kim and Lee [8] proposed Deep Image Quality Assessment (DeepQA) where the behavior of HVS is analyzed from the data distribution of IQA datasets. The sensitivity maps were evaluated for various distortion types and degrees of distortion. Subjective score requires reference images.

Y. Li et al. [9] proposed SESANIA where shearlet transform and deep neural networks (stacked autoencoders) is used instead of conventional regression machines. This framework is enhanced to calculate the quality of image in local regions. Liu, Weijer, and Bagdanov [10] used Siamese Network for ranking images in order of image quality. The relative image quality is known for which synthetically generated distortions are used. This helps to solve the issue of the limited size of the IQA dataset. These ranking image sets can be constructed automatically without the requirement of painful effort of labeling by human. This technique uses synthetic images. Saad et al. [11] introduced a Natural Scene Statistics (NSS) based methodology which uses discrete cosine transform (DCT) technique. This method was based on a Bayesian technique to evaluate the image quality scores when features retrieved from the image is given.

Kede Ma et al. [12] proposed an optimized neural network for assessing blind image quality. First, distortion is identified and then the quality prediction is done using the features obtained during distortion identification.

Fei Gao et al. [13] proposed Deep Similarity for image quality assessment (Deep Sim) framework. First, the features of the original and tested images are received from Image Net pretrained VGGNet without any further training. Then, the local similarities between the features of those corresponding images are calculated. At last, the local quality indices are eventually pooled altogether to evaluate the quality index.

Min et al. [14] proposed the concept of multiple pseudo reference images, which are generated from distorted images by applying various levels of distortion. As a result, the quality of a pseudo reference image (PRI) is generally lower than that of its distorted counterpart. The idea behind this methodology is to generate a series of PRI by further degrading the distorted image, and then use local binary patterns (LBP) to calculate the similarity between them to evaluate its quality.

Talebi and Milanfar [15] proposed a convolutional neural network based methodology known as NIMA which is used to predict the distribution of human opinion scores. The network may be used to score images in a way that closely resembles human perception. Its goal is to forecast image technical and aesthetic attributes.

Hou et al. [16] proposed a blind IQA that directly learns qualitative evaluation and predicts scalar values for general usage and fair comparison. Here, the natural scene statistics features are used to represent the images. A discriminative model is trained to distinguish the characteristics into five ranks, that correlate with five rational notion, i.e., bad, poor, fair, good and excellent.

Bose et al. [17] proposed a neural network based method for IQA that enables feature learning and regression in an end-to-end framework. A siamese network using CNN is used with both original and distorted images as input for FR-IQA whereas one branch of siamese network is discarded where the distorted image is used as input for NR-IQA. It incorporates a weighted average patch aggregation that implements a method for pooling local patch qualities to global image quality.

Based on selected feature similarity and ensemble learning, Hammou et al. [18] suggested an ensemble of gradient boosting (EGB) measure. To characterise the perceptual quality distance between the pristine and distorted/processed images, the features obtained from various layers of deep CNN are analyzed. Kang et al. [19] proposed a compact CNN for calculating image quality and identifying distortions. The parameter reduction at the fully connected layers makes this model less prone to overfitting.

III. MOTIVATION

The main motivation behind image quality assessment is to quantify visual perception of humans for image quality so that quality evaluation of images can be done. Digital images intend to degrade during the process from generation to consumption. Different kind of distortions are introduced in the process of transmission, post processing, or compression of images such as white noise, Gaussian blur, or impeding artifacts. This affects the visual experience of users while seeing image content on various online websites. A dependable IQA algorithm can assist in quantifying the quality of images acquired from the web and also helps to measure the performance of image processing algorithms precisely, such as image-compression and super-resolution, from the point view of a human.

a) Drawbacks of Using CNNs to NR-IQA

Because of its high representation capability and improved performance, convolutional neural networks are the most popular type of neural networks for working with image data. The quantity of the training dataset has a major impact on the performance of neural networks. However, compared to the most frequent computer vision dataset, the currently available IQA datasets are substantially smaller. In contrast to classification datasets, IQA datasets necessitate a time-consuming and sophisticated psychometric experiment. Various data augmentation techniques, such as horizontal reflection, rotation, and cropping, can be employed to enhance the size of the training dataset. The human visual system's (HVS) perception process is made up of several complex processes. It makes training a deep learning model more difficult with a limited dataset. The visual sensitivity of the HVS

changes with the spatial frequency of stimuli, and texture prevents concurrent picture alterations.

b) Applications of IQA

IQA has a diverse variety of computer vision and image processing usage. For example:

- For quantization, an image compression algorithm can use quality as an optimization parameter.
- Image transmission systems can be created to assess quality and distribute different streaming resources accordingly.
- Image recommendation algorithms can be created to rank photos according to perceptual image quality.
- Depending on the image quality desired, several device characteristics for digital cameras can be modified.

IV. PROBLEM STATEMENT

Image Quality Assessment is different from other image processing applications. Unlike segmentation, object detection or classification, preparing IQA dataset is time-consuming and requires complicated psychometric experiments. Therefore, the generation of huge datasets is costly because it requires the supervision of experts which are responsible of ensuring the correct implementation of the experiments. The next drawback is that data augmentation is not preferred because the pixel structure of original images must not be changed. In this paper, an image quality assessment model is developed to calculate the quality of blind images. The distorted images and their ground-truth subjective scores are used for training the CNN model.

V. METHODOLOGY

a) Image Normalization

Image normalization is required because it ensures that the data distribution of each input pixel in the image is consistent. This aids in convergence while doing the training of the neural network. The mean is subtracted from each pixel value, and the result is divided by the standard deviation. Such data would be distributed in a Gaussian distribution centered at zero. The pixel numbers for image input must be positive. As a result, the normalized data must be scaled in the range [0,1] or [0,255]. First, preprocessing is done where the input images are transformed into grayscale, and then they are reduced from their low-pass filtered images. The low-frequency image is retrieved by downscaling the input image to 1/4 and upscaling it again to the original image size. A Gaussian low-pass filter along with subsampling was used to resize the images. The reasons for this kind of normalization is that image distortion doesn't affect the low-frequency component in images. For instance, GB removes high-frequency details, white noise (WN) introduces random

high-frequency components to images, and blocking artifacts introduces high-frequency edges. The distortions caused by JPEG is due to excessive image compression. The human visual sensitivity (HVS) is not sensitive to a change in the low-frequency component of the image. The sensitivity reduces rapidly at low frequency.

There is the possibility of losing information while applying a normalization scheme.

b) Architecture

Model for Non-Screen Content IQA: Here a blind image quality assessment method based on CNN is proposed. The features from the CNN are used for a final quality prediction. The design of the network resembles the design of VGG-16 network. The architecture of CNN for synthetic distortion is shown below in figure 1. The existing dataset consists of a subjective score for each distorted image. The model is fine-tuned to evaluate the subjective scores once the training of neural network is completed with enough training data set. The proposed model is fine-tuned on target subject-specific datasets using a variation of stochastic gradient descent.

The kernel size of the convolutions is 3 x 3. A kernel size of two is used in order to diminish the spatial density in both directions by half. The nonlinear activation function ReLU is used. The feature activations of the final convolution layer's are averaged globally across spatial locations. At the end of the network, three fully connected layers and the ReLU layer are added.



Fig. 1: Synthetic CNN

Model for Screen Content IQA: Here a model based on neural network for screen content image quality assessment called SCIQA is used. The SCI CNN architecture is shown in figure 2. It consists of 8 convolution layers, 4 max-pooling layers, and 2 fully connected layers. All convolution layers have a filter size of 3 x 3 with stride of 1 pixel. A 2 x 2 pixel kernel with stride of 2 pixels is used in each pooling layer. Each convolutional layer's boundary is padded with zeros to improve network speed.

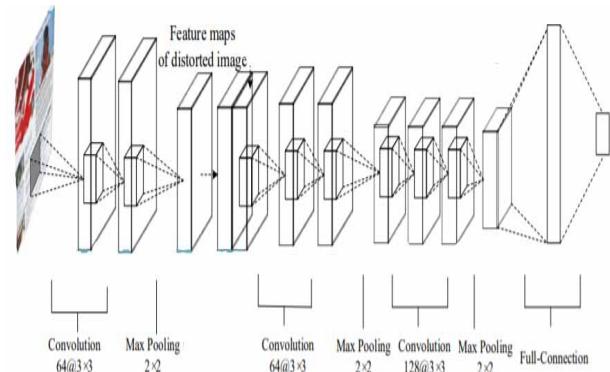


Fig. 2: SCIQA Model

c) Subjective Score

After the model has been trained, it is used to predict subjective scores for the distorted image. As illustrated in Fig. 2, the trained network is connected to a global average pooling layer before the fully connected layers. A 128-dimensional feature vector is created by averaging the feature map over the spatial domain. The adaptive moment estimation optimizer (ADAM) was used to change the normal stochastic gradient descent approach for better optimization convergence.

VI. EXPERIMENT RESULTS AND ANALYSIS

a) Hardware and Software

The experiments has been conducted, and the results were obtained with a laptop with Intel Processor, 8 GB RAM, and 512 GB SSD. As for software, we have used Python as the programming language, and the libraries such as TensorFlow, Keras, SciPy, Matplotlib, etc. in the Jupyter Notebook. The input pipeline for the model is created using TFDS API.

b) IQA Dataset

The IQA datasets consists of distorted images along with their corresponding pristine images. It also have subjective quality scores for distorted images which is obtained after conducting a psychometric experiments using human subjects. Human opinions are taken for these distorted images with reference to pristine images using some pre-defined range for quality measurement. Various IQA datasets were utilized to measure the performance of the proposed algorithm: LIVE IQA dataset, LIVE multiply distorted (LIVE MD) dataset, and UniMiB MD-IVL dataset. The summary of datasets is given in Table I.

- The LIVE IQA dataset consists of following types of distortion: WN, JP2K compression, GB, and Rayleigh fast-fading channel distortion [20][21][22].
- The LIVE MD dataset consists of two categories of images based on distortion combinations applied. First category has images distorted by GB along with JPEG and the second category has images distorted by combination of WN and GB [23].

- The IVL dataset is generated from 10 reference images which is selected from various samples both in terms of low-level features (frequencies, colors) and high level features [24]. This dataset consists of multiple distorted images with 400 images distorted by noise and JPEG distortions.

Cardinal rating is provided by human observer for all distorted images corresponding to their reference images in the dataset from a pre-defined scale which is considered as Mean Opinion Score (MOS). Hence, each distorted image in the dataset has a corresponding ground-truth subjective quality score.

Table 1: Summary of IQA Datasets Used

| Dataset | References | Distortion | Total Samples |
|----------|------------|------------|---------------|
| LIVE IQA | 29 | 5 | 982 |
| LIVE MD | 15 | 2 | 450 |
| MD-IVL | 10 | 2 | 400 |

c) Evaluation Metrics

Unlike traditional pixel-based metrics like PSNR, SSIM, etc. which were used in the past for evaluating IQA algorithms, here the evaluation of the IQA algorithm is done using two statistical measures: SROCC and PLCC i.e., Spearman’s rank- order correlation coefficient and Pearson’s linear correlation coefficient respectively. The PLCC is calculated using the following formula:

$$PLCC = \frac{\sum_{i=1}^n (\hat{S}_i - \mu_{\hat{S}})(S_i - \mu_S)}{\sqrt{\sum_{i=1}^n (\hat{S}_i - \mu_{\hat{S}})^2 (S_i - \mu_S)^2}} \quad (1)$$

where \hat{S}_i and S_i are the predicted and ground-truth subjective scores of the i th image, and $\mu_{\hat{S}}$ and μ_S denote the mean of each. The SROCC is calculated using the following formula:

$$SROCC = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

where n denotes the number of images and d_i is the difference between predicted score and ground-truth score of image.

d) Results and Analysis

i. Performance on Individual Distortion Types

There are 5 distortion types in LIVE IQA dataset. The distortion types are Fast Fading (FF), JPEG, Gaussian Blur (GB), JP2K, and White Noise (WN). The PLCC and SROCC values for each individual distortion type is evaluated using the DIQA [25] framework. In Table II the PLCC and SROCC values are compared based on the individual distortion type using DIQA framework. For WN, the PLCC and SROCC values are highest whereas for JPEG, it is the lowest. Since JPEG affects the image less compared to other distortion types, so the highest values are for WN distortion type.

Table 2: Comparison of PLCC and SROCC values for different distortion types on LIVE IQA Dataset using DIQA [25] framework

| Distortion Type | PLCC | SROCC |
|-----------------|--------|--------|
| JPEG | 0.9713 | 0.9551 |
| JP2K | 0.9759 | 0.9686 |
| GB | 0.9767 | 0.9713 |
| WN | 0.9881 | 0.9918 |
| FF | 0.9748 | 0.9622 |

In Table II, the PLCC and SROCC values are compared based on the individual distortion type using DNSSCIQ frame- work.

Table 3: Comparison of PLCC and SROCC values for different distortion types on LIVE IQA Dataset using DNSSCIQ framework

| Distortion Type | PLCC | SROCC |
|-----------------|--------|--------|
| JPEG | 0.9827 | 0.9624 |
| JP2K | 0.9693 | 0.9656 |
| GB | 0.9727 | 0.9697 |
| WN | 0.9881 | 0.9918 |
| FF | 0.9413 | 0.9447 |

Figure 3 shows the comparison of SROCC and PLCC values for various distortion types in the LIVE IQA dataset using DNSSCIQ framework.

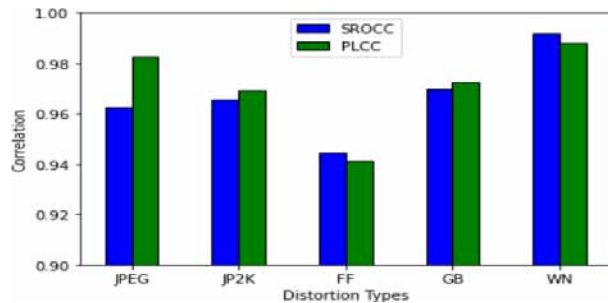


Fig. 3: Comparison of PLCC and SROCC values for various distortion types using DNSSCIQ framework

Table 4: Comparison of PLCC and SROCC values for different model depth on LIVE IQA Dataset

| Model Depth | PLCC | SROCC |
|-------------|--------|--------|
| 5 | 0.9699 | 0.9649 |
| 6 | 0.9769 | 0.9712 |
| 7 | 0.9799 | 0.9752 |
| 8 | 0.9809 | 0.9742 |
| 9 | 0.9767 | 0.9738 |
| 10 | 0.9792 | 0.9730 |

ii. Effect of Model Depth

To determine the influence of model depth, six models with different numbers of convolution layers of DIQA [25] was used.

Convolution layers 1 to 4 and convolution layer 8 was used for the shortest setting. After the Conv6 layer, two 3×3 convolution layers with 64 filters were appended in the longest setting. Figure 4 shows the

accuracy comparison among the models on the LIVE IQA dataset.

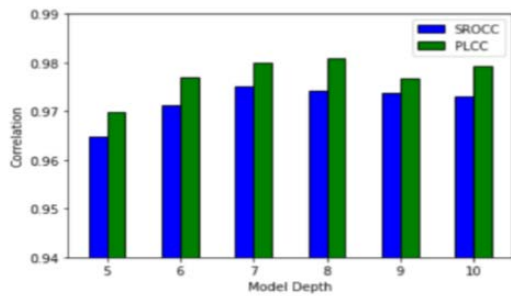


Fig. 4: Comparison of PLCC and SROCC values according to model depth

Table III shows the PLCC and SROCC values for different model depth. When the depth was 5, the PLCC and SROCC values were the lowest. When the depth is increased, the correlation coefficient got saturated around 0.97. This may cause overfitting when more convolution layers are used. Hence, it is concluded that the 8 convolutional layers are good enough for the proposed framework.

iii. Performance on Individual Datasets

The different datasets are used for evaluating the proposed algorithm. The evaluation metrics such as PLCC and SROCC are used. The datasets are having various types of distortions. In some datasets, various distortion types are combine to produce the distorted image. The DIQA method is evaluated on three different IQA dataset individually. The datasets used are LIVE IQA, LIVE MD and MD IVL. Table V shows the comparison of PLCC and SROCC values for individual datasets using DIQA method. For LIVE IQA dataset, the PLCC and SROCC values are highest

The DNSSCIQ method is evaluated on three different IQA dataset individually. The datasets used are LIVE IQA, LIVE MD and MD IVL. Table VI shows the comparison of PLCC and SROCC values for individual datasets using DNSSCIQ

Table 5: Comparison of PLCC and SROCC values for different IQA Datasets using DIQA framework.

| Dataset | PLCC | SROCC |
|----------|--------|--------|
| LIVE IQA | 0.9809 | 0.9742 |
| LIVE MD | 0.9545 | 0.9561 |
| MD IVL | 0.9622 | 0.9617 |

Table 6: Comparison of PLCC and SRCC values for different IQA Datasets using DNSSCIQ

| Dataset | PLCC | SRCC |
|----------|--------|--------|
| LIVE IQA | 0.9867 | 0.9799 |
| LIVE MD | 0.9656 | 0.9685 |
| MD IVL | 0.9696 | 0.9702 |

The PLCC and SROCC values are compared for various IQA datasets like LIVE, LIVE MD and MD IVL in figure 5.

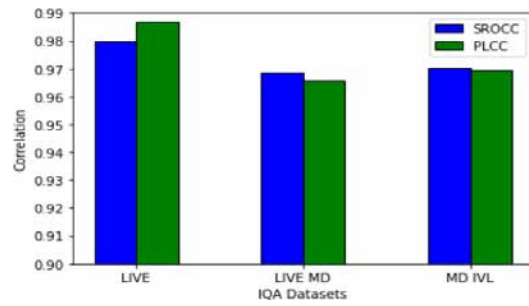


Fig. 5: Comparison of PLCC and SROCC values for various IQA datasets using DNSSCIQ framework

iv. Reliability Map

To find the effect of reliability map, the outputs of various configuration is shown in Table VII. It shows that there is an improvement in performance when reliability map is used. Reliability map helps to create homogeneity across the image irrespective of low-frequency components or high-frequency components in the distorted image. This provides the information about the importance of reliability map.

Table 7: Comparison of PLCC and SROCC values with and without Reliability Maps

| Reliability Map | PLCC | SROCC |
|-----------------|--------|--------|
| w/o | 0.9545 | 0.9561 |
| w | 0.9809 | 0.9742 |

v. NR-IQA Methods

In Table VIII, the PLCC and SROCC metrics of different methods are compared. The different methods are Deep CNN Based Blind Image Quality Predictor (DIQA) [25], Synthetic Convolutional Neural Network (S-CNN) and Screen Content Image Quality Assessment

Table 8: Comparison of PLCC and SROCC values for different method on LIVE IQA Dataset (SCIQA). The S-CNN is having highest PLCC and SROCC values.

| Method | PLCC | SROCC |
|--------|--------|--------|
| DIQA | 0.9809 | 0.9742 |
| S-CNN | 0.9867 | 0.9799 |
| SCIQA | 0.9338 | 0.9229 |

Figure 6 shows the reference image on the left and distorted image with gaussian blur on the right. The image is obtained from LIVE IQA dataset.



Fig. 6: Reference Image on left and Distorted Image (Gaussian Blur) on right

Figure 7 shows the reference image, distorted image in grayscale, error map, reliability map, perceptual error map and sensitivity map in this order. The image is obtained from LIVE IQA dataset.

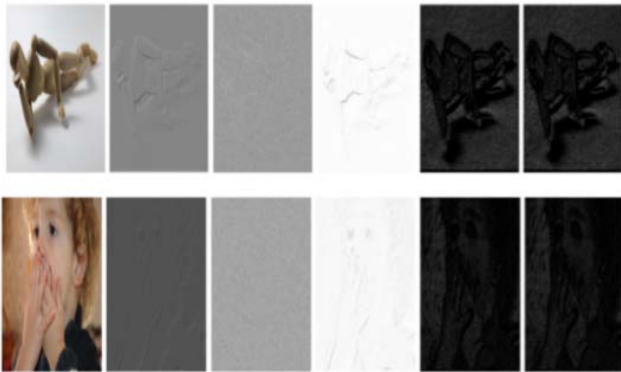


Fig. 7: Reference Image, Distorted Image (Gray Scale), Error Map, Reliability Map, Perceptual Error Map, and Sensitivity Map

vi. Correlation Plot

Correlation plot shows the correlation between any numerical variables. The correlation coefficient is calculated to determine the correlation between two variables.

Figure 8 shows the correlation plot of ground truth and predicted subjective scores. The ground truth scores are provided in the dataset for each distorted image and DNSSCIQ

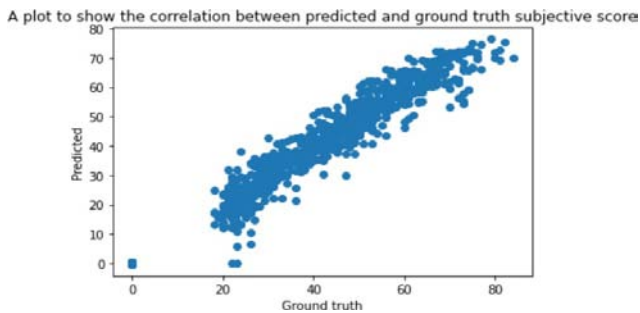


Fig. 8: Correlation Plot

framework is used to obtain the predicted subjective score. From the plot, it is concluded that DNSSCIQ is able to calculate the subjective scores almost close to ground-truth values.

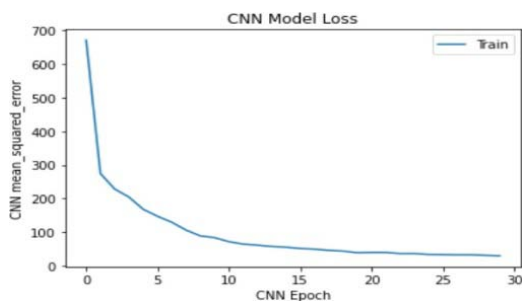


Fig. 9: Loss vs. Epoch Graph

vii. Loss Graph

Figure 9 shows loss vs. epoch graph. Here mean squared error is used as loss function. The loss decreases as the number of epochs increases during training. The performance of the model improves with decrease in the loss.

VII. CONCLUSION

A deep CNN-based approach for Non-Screen Content and Screen Content IQA called DNSSCIQ is proposed. In the DNSSCIQ, the input normalization for the distorted images are done first. Then, the distorted image along with its ground-truth subjective score is provided to the neural network for training to obtain more meaningful feature maps. Once the training is completed, the feature maps are globally average pooled and fed the fully connected layers to get the final subjective score of the distorted image. The performance of the DNSSCIQ is good irrespective of the dataset selected is shown by using various datasets from different sources for training and final quality prediction. In addition to this, distortion-specific evaluation of different datasets is done and the output is compared.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Y. Li et al., "No-reference image quality assessment with shearlet transform and deep neural networks," *Neurocomputing*, vol. 154, pp. 94–109, Apr.2015.
2. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
3. C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793–799, May 2011.
4. A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350– 3364, Dec. 2011.
5. [5]H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 305–312.
6. [6]J. Xu, P. Ye, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation", *IEEE Transactions on Image Processing*, vol. 25, no. 9, Sept. 2016.
7. Qiaohong Li, Weisi Lin, Jingtao Xu and Yuming Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. 18, no 12, Dec. 2016.
8. Kim and Lee, "Deep Learning of Human Visual Sensitivity in Image Quality Assessment Frame-

- work," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
9. Y. Li et al, "No-reference image quality assessment with shearlet transform and deep neural networks," *Neurocomputing*, vol 154, April 2015.
 10. Xialei Liu, Joost Van de Weijer and Andrew D. Bagdanov, "RankIQA: Learning from Rankings for No-reference Image Quality Assessment," IEEE International Conference on Computer Vision (ICCV), Dec 2017.
 11. M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
 12. K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang and W. Zuo, "End-to-End Blind Image Quality Assessment Using Deep Neural Networks," in *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202-1213, March 2018.
 13. Fei Gao, Yi Wang, Panpeng Li, Min Tan, Jun Yu, Yani Zhu, DeepSim: Deep similarity for image quality assessment, *Neurocomputing*, Volume 257, 2017.
 14. X. Min, G. Zhai, K. Gu, Y. Liu and X. Yang, "Blind Image Quality Estimation via Distortion Aggravation," in *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 508-517, June 2018, doi: 10.1109/TBC.2018.2816783.
 15. Hossein Talebi and Peyman Milanfar, "NIMA: Neural Image Assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998- 4011, 2018.
 16. W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.
 17. S. Bosse, D. Maniry, K. M u"ller, T. Wiegand and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," in *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206-219, Jan. 2018, doi: 10.1109/TIP.2017.2760518.
 18. D. Hammou, S. A. Fezza and W. Hamidouche, "EGB: Image Quality Assessment based on Ensemble of Gradient Boosting," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 541-549, doi:10.1109/CVPRW53098.2021.00066.
 19. L. Kang, P. Ye, Y. Li and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 2791-2795, doi: 10.1109/ICIP.2015.7351311.
 20. H.R. Sheikh, Z.Wang, L.Cormack and A.C. Bovik, "LIVE Image Quality Assessment Database Release 2", <http://live.ece.utexas.edu/research/quality>.
 21. H.R. Sheikh, M.F. Sabir and A.C. Bovik," A statistical evaluation of recent full reference image quality assessment algorithms", *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.
 22. Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol.13, no.4, pp. 600-612, April 2004.
 23. Dinesh Jayaraman, Anish Mittal, Anush K. Moorthy and Alan C. Bovik, Objective Quality Assessment of Multiply Distorted Images, *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2012.
 24. <http://www.ivl.disco.unimib.it/activities/imagequality/>
 25. J. Kim, A. Nguyen and S. Lee, "Deep CNN-Based Blind Image Quality Predictor," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 11-24, Jan. 2019.



Acoustic Features based Accent Classification of Kashmiri Language using Deep Learning

By Shehzen Sidiq Malla

Abstract- Automatic identification of accents is important in today's world, where we are surrounded by ASR systems. Accent classification is the problem of knowing the native place of a person from the way He/She speaks the language into consideration. Accents are present in almost all the languages and it forms an important part of the language. Accents are produced from prosodic and articulation characteristics; in this research the aim is to classify accents of Kashmir Language. We have considered using the MFCC and Mel spectrograms for our research. A lot of research has been done for languages like English and is being done in this field and many models of machine learning and deep learning have shown state of the art results, but this problem is new for Kashmiri Language. The accents in Kashmir, vary from area to area and we have chosen 6 areas as our classes. We extracted the features from the audio data, converted those features into Images and then used the CNN architectures as our model. This research can be taken as base research for further researches in this language. Our custom models achieved the loss of 0.12 and accuracy of 98.66% on test data using Mel spectrograms, which is our best for our features.

Keywords: *accent classification, CNN, RELU, mel-spectrograms, MFCC.*

GJCST-D Classification: *1.2.7*



Strictly as per the compliance and regulations of:



Acoustic Features based Accent Classification of Kashmiri Language using Deep Learning

Shehzen Sidiq Malla

Abstract- Automatic identification of accents is important in today's world, where we are surrounded by ASR systems. Accent classification is the problem of knowing the native place of a person from the way He/She speaks the language into consideration. Accents are present in almost all the languages and it forms an important part of the language. Accents are produced from prosodic and articulation characteristics; in this research the aim is to classify accents of Kashmir Language. We have considered using the MFCC and Mel spectrograms for our research. A lot of research has been done for languages like English and is being done in this field and many models of machine learning and deep learning have shown state of the art results, but this problem is new for Kashmiri Language. The accents in Kashmir, vary from area to area and we have chosen 6 areas as our classes. We extracted the features from the audio data, converted those features into Images and then used the CNN architectures as our model. This research can be taken as base research for further researches in this language. Our custom models achieved the loss of 0.12 and accuracy of 98.66% on test data using Mel spectrograms, which is our best for our features.

Keywords: *accent classification, CNN, RELU, mel-spectrograms, MFCC.*

I. INTRODUCTION

Kashmiri or Koshur is a Dardic language subgroup from Indo-Aryan, spoken by over seven million Kashmiris [Wikipedia]. There are many accents Spoken in Kashmir. There are some major accents and some minor accents in this language. This leads to diversity in the language and adds to its beautiful sounds and variations. The aim of this research is to classify these different accents. Although many accents are being spoken in this language, for this research, we have classified the prominent accents belonging to Kupwara, Srinagar, Islamabad, Shopian, and Bandipora. The proposed approach is on the basis of using Convolution Neural Networks (CNN) and training Neural networks on the images of features extracted from the audio files. The features are Mel- spectrogram and MFCCs. Our approach uses CNN as the classifier and MFCCs, Mel spectrograms as Features. Three types of MFCCs are extracted, 13, 24 and 36. We got excellent results on our dataset.

Accent classification refers to the problem of inferring the native language of a speaker from his or her foreign accented speech. Identifying idiosyncratic differences in speech production is important for

improving the robustness of existing speech analysis systems. For example, automatic speech recognition (ASR) systems exhibit lower performance when evaluated on foreign accented speech. By developing pre-processing algorithms that identify the accent, these systems can be modified to customize the recognition algorithm to the particular accent [1] [2]. In addition to ASR applications, accent identification is also useful for forensic speaker profiling by identifying the speaker's regional origin and ethnicity in applications involving targeted marketing [3] [4]. In this paper we propose a method for classification of 11 accents directly from the speech acoustics.

For example, Deshpande et al. used GMMs based on formant frequency features to discriminate between standard American English and Indian accented English [6]. Chen et al. explored the effect of the number of components in GMMs on classification performance [7]. Tang and Ghorbani compared the performance of HMMs with Support Vector Machine (SVM) for accent classification [8]. Kumpf and King proposed to use linear discriminant analysis (LDA) for identification of three accents in Australian English [9].

Artificial neural networks, especially Deep Neural Networks (DNNs) and Recurrent Neural Networks (RNNs) and CNNs have been widely used in state-of-the-art speech systems and Image Processing Systems [10] [11] [12] [13]; however, in the area of accent identification, there are only a few studies evaluating the performance of neural networks [14] [15]. Nonetheless, in a related area, language identification (LID), neural networks have been investigated exhaustively [16] [17] [18]. In a recent paper [19], where they used spectrograms for accent classification and speaker recognition and achieved an accuracy of 92%. Inspired by their work, we also propose to use Mel-Spectrograms and MFCCs for our research

The rest of the paper is organized as follows: In section 2, we discuss the collection and making of dataset. In section 3, we discuss proposed system and discuss in detail the features that we have used for our research. In section 4, we discuss the experiments and show our results and finally in section 5, we conclude our research.

Author: e-mail: mallashehzen786@gmail.com

II. DATASET

a) Collection of Dataset

Data is very important in every machine learning and deep learning project or deep learning research. The data, we required for our research was, the audio files of people speaking some sentences that we choose. These sentences, to some extent captured a wide range of accent changes in the spoken Kashmiri Language. In total, 20 sentences were chosen for research purposes and people were recorded speaking these sentences in their native accents of Kashmir language. The data was collected from 5 districts or areas of Kashmir and all these files were saved with the extension of 'ogg', which in preprocessing, were changed to 'wav' format. In total, we got almost 100 voice samples from each area and thus we had, 500 total voice samples of these sentences spoken by different people.

b) Making of dataset

The data we had, were audio files and we decided on getting the MFCC and Mel-Spectrograms from these audio files. So, our final dataset consisted of images of MFCC and Mel-Spectrograms. Since deep learning models require huge amount of data, we had to augment the data to increase the size of our dataset. There are many great techniques of augmenting the data, when it comes to images and audio. Since the images were of features, we could not use the normal augmentation techniques like distortion, rotation and many more [reference]. A special kind of augmentation known as specAugment [20], which produces augmented images on spectrograms was used. This augmentation performed following operations on the Images of Mel-Spectrograms. 1) Frequency masking is where certain part of the frequency is masked out, and 2) Time masking, where certain part of time is masked out. Even though, we performed augmentation on Mel-Spectrogram images, the data was not enough. so, we had to perform the augmentation on the audio files also. The audio files were augmented by increasing the speed, pitch and amplitude of the audio files, thus giving somewhat variability in the initial dataset of audio files.

After performing, such augmentations we had large sufficient dataset for deep learning.

III. PROPOSED SYSTEM

a) Architecture

We used CNN based architecture with ReLU activation function for internal nodes and SoftMax function to output the probability distribution of output classes. CNN [19] models show state of the art performance with image data. Since our motive was to extract the features from the audio data and plot them as images and then those images were input to the model, so we chose the model based on CNN

architecture. Our model has, six convolution layers and six Maxpooling layers, with 5 dense layers and a flatten layer.

[image of model] Accuracy and loss varied based on the feature used and learning rate of the model. We choose different learning rates based on different features that were input to the model. Our models were trained on learning rates between 0.001 to 0.0001

b) Features Used

Many features have been used in researches of audio processing. We decided to keep our research simple so we decided on two features, MFCC and Mel Spectrograms. MFCC have been found to perform well in case audio classification [21] purposes and Mel spectrograms and Spectrograms have also shown such performance in many cases [19]. Different number of coefficients can be used, mostly 13 are taken. The selection of such number of constants, depends on the problem in hand. We experimented on various number of coefficients and finally decided on 13, 24 and 36 coefficients. These features were extracted and plotted as images and then such images were input to our model.

Mel Spectrograms - A Mel spectrogram is a spectrogram that converts the frequencies to the Mel scale. When the spectrogram from the audio file is plotted using Mel scale, we get the Mel-Spectrogram. These spectrograms were plotted as images, same as the MFCCs and given input to the model.

All these operations of feature extraction were done using librosa library [22], which makes working with audio very easy.

IV. EXPERIMENTAL SETUP AND RESULTS

Different experiments were performed on different features and different learning rates were set during the training of the models.

The features, were stored in two ways -

Images of the features were generated, and in other features were extracted and a dimension was added, no image was generated and the features were stored in JSON data format. Below, we show the results of various experiments:

a) Experiment 1

This Experiment was done using images of Mel spectrograms and MFCC for the CNN having 3 color channels. In this experiment, the images were generated from the audio files. Those images were saved and later were loaded back into the model. The models were trained on those images and evaluated on the validation and testing sets.

i. Mel Spectrograms

The below figure shows the metrics graphically and we can conclude from the graph that the model is showing state of art results on our data.

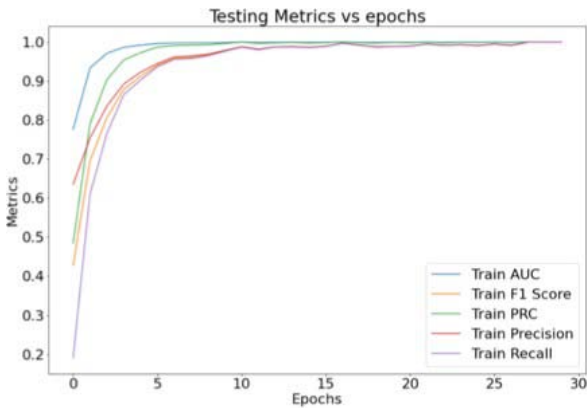


Fig. 1: Plot of Metrics of training data

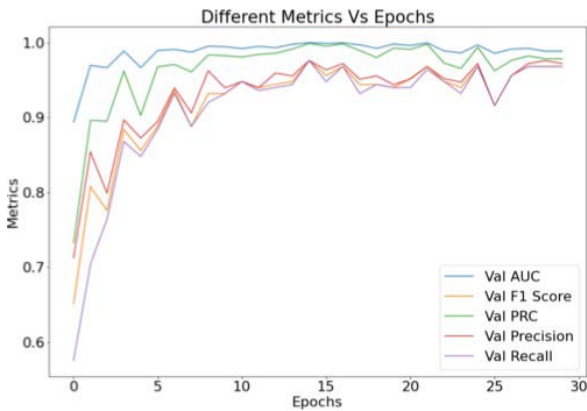


Fig. 2: Plot of metrics of validation data.

The above figure shows the metrics for the validation set. All these metrics were evaluated for the model which uses spectrogram images as the input.

Loss - 0.1259424239397049
Accuracy - 0.9866666793823242
F1 Score - 0.9866666793823242
Precision - 0.9866666793823242
Recall - 0.9866666793823242
AUC - 0.9916666746139526
PRC - 0.9788309931755066

Fig. 3: Metric Scores for testing data.

These above results were calculated on the testing data and we can conclude that our model performed much better than expected and showed state of the art performance on our data.

ii. MFCCs

The MFCC features were extracted from the audio files and plotted as images and these Images were saved and loaded at the time of model training. The following figures show the accuracies and losses with respect to the epochs. Three types of constants were extracted and same model was trained on these images generated from the audio files. The training was done using the training data, and validated on validated. Following table shows our results.

Table 1: Validation losses and Validation accuracies for Different features [images]

| Feature | Validation Loss | Validation Accuracy |
|------------------|-----------------|---------------------|
| Mel Spectrograms | 0.0392 | 0.9848 |
| MFCC 13 | 0.04 | 0.98 |
| MFCC 24 | 0.031 | 0.99 |
| MFCC 36 | 0.06 | 0.97 |

From the above table, we can see that the 24 constants performed slightly better than the others on validation data.

Using the images as input to the model, Mel-Spectrograms and MFCC 24 constant features performed better than the other features.

b) Experiment 2

This experiment was done using Json files of extracted features and giving to CNN having 1 color channel.

In this experiment, the features were extracted and were saved in JSON files. No Images were generated for this data. Then the features were loaded back and the model was trained on these features. The below table show the Testing accuracies and testing Losses for various features extracted from audio files.

Table 2: Testing Loss and Testing Accuracy for Various Features [JSON]

| Feature | Testing Loss | Testing Accuracy |
|---------|--------------|------------------|
| MFCC 13 | 0.086 | 0.87 |
| MFCC 24 | 0.110 | 0.87 |
| MFCC 36 | 0.507 | 0.865 |

c) Experiment 3

This experiment was done by splitting the audio files in the chunks of 2 seconds.

In this experiment the audio was splitted into two second chunks a The Mel spectrograms features were extracted from the splitted audio files and then saved as images and as well as JSON files. Following accuracies and Losses were calculated on validation data.

Table 3: Validation Loss and Accuracy for 2 Sec Splitted Data

| Type of Feature | Validation Loss | Validation Accuracy |
|-----------------|-----------------|---------------------|
| Images | 0.0331 | 0.98 |
| JSON files | 0.069 | 0.97 |

V. CONCLUSION

This research paper proposes a solution to the accent classification for Kashmiri Language using Convolutional Neural networks. The solution is based on deep learning techniques using CNNs that adapt to the multi-dimensional data. CNNs provide solution to this

problem using the supervised approach where they first undergo training session during which they are fed with labelled data from which they learn the relationships in the data and attain the learning capability. In later stage, they are presented with unseen data of same domain and are able to make remarkable inferences from this unseen data by utilizing the attained learning capability. In addition to reporting the state-of-art classification results, its accuracy is also remarkable. In our research, we can conclude that the models and the data that we used, the Mel Spectrograms performed better and showed better performance than the MFCCs, also we saw that the images with 3 color channels performed better than the features that were saved with extended dimension. Overall we can conclude, our model showed state of the art performance for the Accent classification of Kashmiri Language with five output classes.

VI. FUTURE IMPROVEMENTS

There is a lot of improvement to be done in this field of research. Since, this is the first research in this language, as we could not find any other research related to Kashmiri language, so the area of improvement is vast. We propose following enhancements for this research-

- Collection of more data for efficient model training
 - a. The dataset can be increased in size
 - b. The dataset can be in such a way, that it captures the maximum of the features and variations present in the language.
- The model can be made more complex and more sophisticated that would be able to handle more data and not underfit.
- Making efficient model for being able to capture most of the features of accent classification
- Improving the classification error and thus being able to classify wide range of the language.
- Making use of different architectures and techniques available for making the overall application most fruitful.
- The classification classes can be increased to more than 5 accents or regions.

REFERENCES RÉFÉRENCES REFERENCIAS

1. L. Kat and P. Fung, "Fast accent identification and accented speech recognition," in Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on, vol. 1. Phoenix, AZ, USA: IEEE, 1999, pp. 221–224.
2. C. Huang, T. Chen, and E. Chang, "Accent issues in large vocabulary continuous speech recognition," *International Journal of Speech Technology*, vol. 7, no. 2-3, pp. 141–153, 2004.
3. D. C. Tanner and M. E. Tanner, Forensic aspects of speech patterns: voice prints, speaker profiling, lie and intoxication detection. Lawyers & Judges Publishing Company, 2004.
4. F. Biadsy, J. B. Hirschberg, and D. P. Ellis, "Dialect and accent recognition using phonetic-segmentation supervectors," 2011.
5. S. Deshpande, S. Chikkerur, and V. Govindaraju, "Accent classification in speech," in Automatic Identification Advanced Technologies, Fourth IEEE Workshop on. Buffalo, NY, USA: IEEE, 2005, pp. 139–143.
6. T. Chen, C. Huang, E. Chang, and J. Wang, "Automatic accent identification using gaussian mixture models," in Automatic Speech Recognition and Understanding, IEEE Workshop on. Madonna di Campiglio, Italy: IEEE, 2001, pp. 343–346.
7. H. Tang and A. A. Ghorbani, "Accent classification using support vector machine and hidden markov model," in Advances in Artificial Intelligence. Springer, 2003, pp. 629–631.
8. K. Kumpf and R. W. King, "Foreign speaker accent classification using phoneme-dependent accent discrimination models and comparisons with human perception benchmarks," in Proc. EuroSpeech, vol. 4, pp. 2323–2326, 1997.
9. G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *Signal Processing Magazine, IEEE*, vol. 29, no. 6, pp. 82–97, 2012.
10. H. Zen and H. Sak, "Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis," in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. Brisbane, Australia: IEEE, 2015, pp. 4470–4474.
11. Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "An experimental study on speech enhancement based on deep neural networks," *Signal Processing Letters, IEEE*, vol. 21, no. 1, pp. 65–68, 2014.
12. Y. Jiao, M. Tu, V. Berisha, and J. Liss, "Online speaking rate estimation using recurrent neural networks," in Acoustics, Speech and Signal Processing, IEEE International Conference on. Shanghai, China: IEEE, 2016.
13. M. V. Chan, X. Feng, J. A. Heinen, and R. J. Niederjohn, "Classification of speech accents with neural networks," in Neural Networks, IEEE World Congress on Computational Intelligence., IEEE International Conference on, vol. 7. IEEE, 1994, pp. 4483–4486.
14. Rabiee and S. Setayeshi, "Persian accents identification using an adaptive neural network," in Second International Workshop on Education Technology and Computer Science. Wuhan, China: IEEE, 2010, pp. 7–10.

15. G. Montavon, "Deep learning for spoken language identification," in NIPS Workshop on deep learning for speech recognition and related applications, Whistler, BC, Canada, 2009, pp. 1–4.
16. R. A. Cole, J. W. Inouye, Y. K. Muthusamy, and M. Gopalakrishnan, "Language identification with neural networks: a feasibility study," in Communications, Computers and Signal Processing, IEEE Pacific Rim Conference on. IEEE, 1989, pp. 525– 529.
17. Lopez-Moreno, J. Gonzalez-Dominguez, O. Pichot, D. Martinez, J. Gonzalez-Rodriguez, and P. Moreno, "Automatic language identification using deep neural networks," in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. Florence, Italy: IEEE, 2014, pp. 5337–5341.
18. Y. Zeng, H. Mao, D. Peng, and Z. Yi, 'Spectrogram based multi-task audio classification', *Multimed Tools Appl*, vol. 78, no. 3, pp. 3705–3722, Feb. 2019, doi: 10/gnkrss
19. D. S. Park *et al.*, 'SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition', *Interspeech 2019*, pp. 2613–2617, Sep. 2019, doi: 10/ghbzt4
20. Zhao, Huimin & Xianglin, Huang & Wei, Liu & Yang, Lifang. (2018). Environmental sound classification based on feature fusion.
21. B. McFee *et al.*, 'librosa: Audio and Music Signal Analysis in Python', Austin, Texas, 2015, pp. 18–24. doi: 10/gf4wxc. MATEC Web of Conferences. 173. 03059. 10.1051/mateconf/201817303059.





This page is intentionally left blank



Neural Network Design using a Virtual Reality Platform

By Luigi Bibbò & Francesco Carlo Morabito

DIIES University

Abstract- The evolution of Deep Learning (DL), a subset of machine learning, has made their use very effective in many artificial intelligence (AI) fields. In parallel Virtual Reality is going wide in many applications thanks to the proliferation of cameras in mobile devices and improved processing efficiency. Data visualization in deep learning is a fundamental element for which it can benefit from the advantages offered by the visualization of the VR for the development of the models. In addition, the researchers can widely use the editing of images and videos in the machine learning process to design a convolutional network suitable for image recognition. In this study, we want to demonstrate the usefulness of this approach in collecting data within virtual reality to train and optimize a convolutional neural network used to recognize human activities (HAR).

Keywords: *machine learning, deep learning, neural nets, visualization, virtual reality.*

GJCST-D Classification: *F.1.1*



NEURALNETWORKDESIGNINGAVIRTUALREALITYPLATFORM

Strictly as per the compliance and regulations of:



Neural Network Design using a Virtual Reality Platform

Luigi Bibbò ^α & Francesco Carlo Morabito ^ο

Abstract- The evolution of Deep Learning (DL), a subset of machine learning, has made their use very effective in many artificial intelligence (AI) fields. In parallel Virtual Reality is going wide in many applications thanks to the proliferation of cameras in mobile devices and improved processing efficiency. Data visualization in deep learning is a fundamental element for which it can benefit from the advantages offered by the visualization of the VR for the development of the models. In addition, the researchers can widely use the editing of images and videos in the machine learning process to design a convolutional network suitable for image recognition. In this study, we want to demonstrate the usefulness of this approach in collecting data within virtual reality to train and optimize a convolutional neural network used to recognize human activities (HAR).

Keywords: machine learning, deep learning, neural nets, visualization, virtual reality.

I. INTRODUCTION

Machine Learning (ML) indicates a research area within Artificial Intelligence needs to create systems capable of autonomous learning without being specifically programmed to carry out this task [1]. This step introduces a new programming vision, in which the goal is no longer to feed input to receive an output from the machine but to allow it to make decisions autonomously.

Deep Learning (DL) is a machine learning technique used to build Artificial Intelligence (AI) systems [2]. It is a technology that creates learning models on several levels. Deep learning systems allow a representation of information at various levels in a hierarchical way and build models that exploit large amounts of visual media data. The learning of data takes place through the use of statistical calculation algorithms. The inspiring principle was to reproduce a system of reasoning that is biologically inspired by the human brain, must be able to behave by simulating the functioning of neurons. The human neuron represents the computational engine that is the basis of deep learning that takes place in Artificial Neural Networks (ANN). A neural network reproduces all those processes in the brain during the learning phase and subsequent recognition phase. We design the artificial neural networks (ANN) to perform complex analysis of large amounts of data by passing it through multiple layers of neurons. The peculiar characteristic of neural networks is learning and adapting to their environment. Another property is to build connectivity models based on the

approximation of the error. These elements make identifying similar patterns in the test data [3].

A neural network requires a training phase and an inference phase to operate correctly. During the training phase, the number of neurons and the levels that will comprise the neural network are defined and exposed to the labeled training data. Finally, the network will learn which data is correct and discard the wrong one. Once the network has been trained, it is possible to proceed to the inference phase in which the network itself is called to evaluate new images proposed to it. These steps are part of the neural network design process, which includes:

- 1) *Determination of the function that the network must be able to perform:*
 - a) *Classification*, this function involves sorting images into various classes and grouping them according to common properties, allowing you to determine whether a specific type of object is present in an image or not;
 - b) *Detection and localization*, this function allows you to identify the position and size of an object by identifying the characteristics of the image;
 - c) *Segmentation*, used to identify which pixels of an image belong to the corresponding objects, allows determining the boundaries and areas of the objects in the images.
- 2) *Choice of framework:* we must make this choice based on the applications to be created, the libraries available, or the developers' competence. In addition, there are numerous accessible frameworks such as, e.g., TensorFlow of Google, Caffe2 of Facebook e Open VINO of Intel, and Pytorch: an open-source solution that is now part of Facebook.
- 3) *Loading training data:* this phase and the training are crucial for the proper functioning of the neural networks. While being favored by the ability to generate features as they learn automatically, deep networks still require large amounts of training data to develop a properly functioning model. The amount of data needed depends on the complexity of the domain we are trying to approximate. The choice of the size of the dataset is closely related to the selection of the number of neurons in the neural network. The total complexity is generally not known before starting the training. For this reason, the

Author α: DIIES University "MEDITERRANEA" Reggio Calabria, Italy.
e-mail: luigi.bibbo@unirc.it

Author ο: DICEAM University "MEDITERRANEA" Reggio Calabria, Italy.

neural network training process is iterative. At the end of the training, the network's performance is analyzed: the results of this analysis provide information on whether the available data is sufficient. The data for training can be found already labeled on public archives or purchased. Alternatively, it is necessary to generate a faked data set and label them. Each type of application to be developed requires its own set of data, complete with all the characteristics used for a correct evaluation.

- 4) *Network training and validation*: training is the phase in which a neural network must acquire the knowledge to interpret specific input data. In this phase, it is necessary to provide examples in input/output pairs through which the network must learn to predict the expected output. The network is a mathematical model whose output is regulated by the weights of the signals received by each neuron. During the training phase, these weights are progressively calibrated so that more and more, the network's output in the face of a specific input approaches the desired one. This phase is implemented by comparing the initial classification with what is suggested by learning. We use an algorithm called Backpropagation is used to train neural networks. It compares the result obtained from a network with the output you want to have. We use the difference between the two results to change the weights of the connections between the network layers starting from the output layer. Then we proceed backward by modifying the weights of the hidden layers and finally those of the input layers. To do this, we use a function called cost appropriate to the problem we want to solve. A part of the data prepared is used for the training phase while the data for the test and validation phase are kept separate. Since pre-trained networks are available, we can use them to classify new objects using the transfer learning technique.
- 5) The process of evaluating new images using a neural network to make decisions is called inference. This step can collect additional test data used as training data for future iterations.

There is a wide variety of deep neural networks (DNN). Convolutional neural networks (CNN) or Deep convolutional neural networks (DCNN) are the types most commonly used to identify patterns in images and video. The name "convolutional neural network" indicates that it employs a mathematical operation called convolution in place of general matrix multiplication in at least one of their layers [4]. They have evolved from traditional artificial neural networks, applying the connectivity pattern between neurons that resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a

restricted visual field region known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visible area. CNN's are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks; each neuron in one layer is connected to all neurons in the next layer. Deep convolutional neural networks mainly focus on object detection, image classification, and recommendation systems and are sometimes used for natural language processing. In the absence of CNN to identify the objects in the images, it was necessary to resort to manual feature extraction methods, making mistakes and taking a considerable amount of time. CNN's are characterized by their ability to optimize filters (or Kernel) through machine learning, differentiating themselves from traditional algorithms in which filters must be manually sized.

A feature common to all architectures of deep neural networks, including CNN's, is their high complexity due to many internal parameters. For this reason, numerous techniques have been developed to facilitate the design and understanding of the internal workings of a network [5]. The lack of interpretability and transparency of neural networks and the absence of explanations on the decisions taken can compromise the confidence in their applicability. Moreover, they are endorsed by the large datasets required to train most deep learning models. Therefore, the possibility of understanding the model's functioning allows one to overcome these problems. Finally, we can understand how deep learning models make decisions and what representations they have learned through deep learning visualization. In this way, we regain confidence in the model [6].

II. DATA VISUALIZATION

Visualization of data represents another crucial phase in the design of neural networks. Designers want to visualize deep learning to understand how deep learning models make decisions and what representations they have learned to make the model as reliable as possible [7]. This notion of general understanding of the model is called interpretability or explicability.

If we consider the training process of a DNN, its visualization facilitates this delicate design phase. Therefore, we need to optimize parameters to minimize the loss function through gradient descent when training neural networks. To ensure that the loss decreases over time, we need to monitor the entire training cycle and test the losses over the "epochs."

Many visualization techniques are limited to showing whether a model is improving during its iterations, in the training process, rather than how it is training and why it makes certain decisions. For example, Martin Becker and al. developed a tool based

on *Gradient Descent Analysis* in deep neural network (DNN) training [8], representing a new interactive visualization tool for exploring DNN learning processes. They used a method based on analyzing a new type of training error curve on three levels of detail. Usually, the comparison method between their training error curves is used to evaluate the learning processes resulting from different hyperparameter settings. With this approach, the authors associate three levels of detail to the three stages of Shneiderman's Visual Information Seeking Mantra. In the first stage, the traditional training error curve is the focus of attention. In the zoom and filter stage, we can explore the variation of the training error along each of the descent directions considered during the gradient descent-based DNN training. The third stage allows for an in-depth exploration of methods for adaptive control of the step width along a given descent direction.

Other techniques that focus visualization on the learning process rather than the magnitude of the error are Deep Tracker [9] and Tensor view [10].

Deep Tracker is a visual analysis system designed to analyze CNN training processes' dynamics and identify unusual patterns within the numerous training logs. The authors associated a series of small hierarchical multiples with a hierarchical structure index mechanism to allow you to explore the entire training process from different levels of detail. They also suggest a new cube-style visualization to reveal the complex correlations across multiple types of heterogeneous training data.

Tensor View is a tool that uses Paraview and Matplotlib to study neural networks built on the TensorFlow framework. It allows visualizing the evolution of the neural network parameters, such as the weights of the filters and gradients, providing direct observation of how the neurons update during training. This tool uses visualization to inform researchers about their hyperparameters; to use quantitative information to eliminate unnecessary neurons, accelerating the training processes.

These two tools are very complex, so expert researchers can only use them. Other visualization techniques aim to provide the right amount of information on the evolution of the training process, facilitating the understanding of even less experienced researchers.

Michelle Peters and al. [11] developed, for non-experts, a technique that intuitively visualizes the training process of a neural network to facilitate the understanding of the decisions of a neural network. They use the *dimension reduction method* Uniform Manifold Approximation and Projection (UMAP) [12] to visualize neuronal activity. They generate a visualization with three plots: 1) - a 2D plot of the test data with the input images, 2) - a 2D plot of the training data, 3) - a line plot of the accuracy of both the testing and training

data. They also create a video, generating a frame for every epoch, showing how the neural network improves during the training phase. This approach demonstrates how an ongoing training video offers more information than a static view at the end of the process and which features are helpful for non-experts to view, highlighting that the insights drawn from video analysis can be beneficial for network design.

Various visualization techniques have been incorporated into interactive tools to aid in the DNN design process.

Bock et al. [13] designed a tool for 3D representation of deep learning algorithms for experts and non-experts with an interactive user interface that allows visualiz

Pezzotti and al. [14] developed a progressive visual analysis technique called *Deepeyes*. Multiple linked views are associated with analyzing a network during training and showing how it changes over time. In addition, the system supports researchers in identifying problems, such as unnecessary filters or layers and information not captured by the network.

Kahng et al. [15] developed *ActiVis*, an interactive visualization system for large-scale deep neural networks, to help users understand how a model derives its predictions. Users can better diagnose discrepancies with their neural network by integrating multiple coordinated views: such as an overview of the model architecture computation graph, of how the deep learning neurons, which together make up the network, are activated by user-specified instances or subsets of instances. Finally, it provides a graphic representation of the model architecture; hence the user can perform a localized inspection of activations on each model level.

A classification of CNN viewing techniques distinguishes them into two categories: feature display and attribution [16]. With the feature visualization technique, we study how the network operates to generate an example; attribution studies which part of the example is responsible for attributing the network in a certain way.

One of the methods for feature visualization is activation maximization [17] which can be applied at different levels of a CNN. This technique highlights that a CNN tends to build its understanding of an image hierarchically on many levels. For example, Nguyen and al. have developed a system based on multifaceted feature visualization (MFV) [18], uncovering the different types of features learned by each neuron in deep neural networks. Previous activation maximization techniques constructed images without regard for the multiple facets of a neuron; instead, this method explicitly shows each neuron's various aspects by producing a synthetic visualization of each of the types of images that activate a neuron. The visualization of the faceted features thus provides a more accurate and more complete description of the role of each neuron.

To attribution methods, Chattopadhyay and al. [19] have developed a methodology that starts from the visualization of the neural network's architecture as a Structural Causal Model (SCM) and calculates each feature's causal effect on the output. Their approach differs from other related works to structural learning [20]. The goal is to discern the causal structure in data provided; their goal is to identify the causal influence on the output of a learned function.

Sattarzadeh and al. [21] propose a model based on visualization maps from multiple levels using an attribution-based input sampling technique and aggregating them to achieve a satisfactory and complete explanation. The methodology adopted to interpret the layers of CNN results in a four-step method of description. In the first three stages, information extracted from multiple levels of CNN is represented in the accompanying display maps. These maps are then merged into a single module to form an explanation map in the last step. The proposed solution allows overcoming some problems encountered in models that use input sampling techniques. These methods have shown excellent fidelity in rationally inferring model predictions. However, they exhibit instability indices since their output depends on sampling randomness (RISE) or random initialization to optimize a perturbation mask. Finally, their algorithm called Semantic Input Sampling for Explanation (SISI) replaces the randomized RISE input sampling technique with an attribution-based sampling technique. Finally, it uses feature maps derived from multiple levels.

The Deep Replay open-source Python package was designed to allow you to visualize how a Deep Learning model in Keras is performing or learning at each iteration/epoch.

Virtual reality (VR), thanks to its peculiar characteristics, represents a valid alternative to the visualization of conventional data. VR technology establishes a new form of human-computer interaction, arousing in the user a new type of experience linked to the concept of "presence" [22]. Presence is the sensation of being physically and spatially placed in an environment. With virtual reality, you can realize this feeling of being in another world, the virtual one. VR can generate experience and support knowledge acquisition thanks to this new communication process. However, the sense of presence developed by VR depends significantly on the characteristics of the technology used. It is associated with the perceptual illusion that the user has to interact with the remote environment, such as if it were present. In the interaction with the VR environment, a perceptual illusion is created in the user. The stimulation of the senses produces cognitive and emotional models consistent with the experimenting environment. When in VR, a user will inevitably compare the appearance of virtual objects with real-world objects and judge the level of congruence. The other element

that favors VR over other visualization techniques is the level of "immersion"[23]. The greater immersion allows you to improve the design business, perform tasks more efficiently and with greater understanding. Immersion in virtual reality is the perception of being physically present in a non-physical world. Perception is created by surrounding the user of the VR system with images, sounds, or other stimuli that provide an evocative environment. Finally, the user can move freely and explore places and spaces through the movements of the head and eyes and interact with objects by grabbing and dragging them. In addition, movements from the physical world are transferred to the virtual world with great precision. This aspect is precisely exploited to simulate physical actions. Finally, workflows and activities can be transferred to the virtual model to verify the correct interpretation, improving the system's accuracy.

These characteristics make virtual reality a suitable environment for simulations and e-learning purposes as they can stimulate users' creativity and accelerate the learning process [24]. This technology is widely applied in various fields ranging from entertainment design to education and rehabilitation and psychological therapies [25]. VR technology makes it possible to improve the rehabilitation-psychological program compared to traditional techniques by creating a more natural process, reproducing the characteristics of home living environments. Other elements that characterize this technique are represented by the possibility of dynamically adjusting the difficulty of the exercises concerning the skills acquired and stimulating the patient's multisensory skills. In the context of training, it has proved to be very effective compared to traditional techniques as they can keep the attention high and focus the topics better [26].

Immersion and interactivity are the elements that suggest the choice of VR for CNN as through them; the user can better focus and understand how it works.

The following sections illustrate the main characteristics of CNN and virtual reality technologies.

III. CONVOLUTIONAL NEURAL NETWORK

From an architectural point of view, a convolutional neural network is a multi-layered feedforward neural network composed of an input layer, hidden layers, and an output layer. The hidden layers are typically convolutional, followed by activation and pooling layers. This sequential design allows convolutional neural networks to learn hierarchical features. Convolutional neural networks (CNN) process data through many layers of artificial neurons. This human brain (CNN) process is constituted by a set of different layers that act as extractors of the features of the input images and a fully connected terminal network that acts as a classifier. It has proved to be an effective

solution for image recognition [27]. They are built to analyze images included within certain data sets and classify objects in images within them. CNN is a network composed of several convolutional layers [28]. Each processing layer comprises a convolutional filter, an activation function (ReLU), a pooling function, and a fully connected layer. At the end of each processing step, an input is generated for the next layer. In the convolutional operations, the trained filter set is convolved with input images to extract the specific feature to create the feature map, which becomes the input for the next filter. Finally, the design of a CNN network requires a training period followed by a test phase. During the training phase, the images are labeled and transferred to the subsequent layers to allow the structure to convert from the representation level of the original input to a higher level and more abstract representation to constitute the reference feature maps with which the network must compare the output feature maps. Once the training and test phase of the network has been completed, we will determine the survey's accuracy level.

Each layer comprises three levels: Convolution, ReLU, and Pooling.

- Convolutional Level (CONV) is the main level of the network. Its objective is to identify patterns. They are multiple, designed to identify features present in the initial image. Each layer learns to extract specific parts of the photos placed at its entrance. Multiple layers in cascade combine the features of the previous layers with higher programmed extraction levels.
- Rectified Linear Unit (ReLU) Level is placed after the convolutional level and can cancel negative values obtained in the previous classes.
- Pool level allows identifying if the study characteristic is present in the previous story. The pooling layer obtains images with a particular resolution at the input and returns the same number of pictures with fewer pixels.

The result of the convolution operations is the production of feature maps obtained with the help of filters that are matrices containing proper values for finding specific characteristics in the input images. At the end of the sequence of convolutional layers, there is then the fully connected level (FC) which aims to establish the identifying classes obtained in the previous levels according to a certain probability.

Each category represents a possible answer that the system will most likely choose.

During the recognition phase, the network performs a classification operation to identify which class the input image belongs to, identifying the one with the highest probability.

The values of the filters are initially chosen randomly. They are subsequently improved at each iteration of the training phase.

We estimate that the model's predictions are plausible; in practice, we measure the discrepancy between actual and predicted values. The error is subsequently processed using the stochastic gradient technique. It is a cyclical technique consisting of two phases:

- * backpropagation;
- * updating the gradient value.

After propagating the forward phase during training, the outputs are produced to determine the prediction error compared with the expected ones. This error is used to calculate the gradient of the loss function. The backward propagation phase then sends the error through the network layers and updates the weights using the stochastic gradient descent to improve the network's performance on the activity it is trying to learn [29].

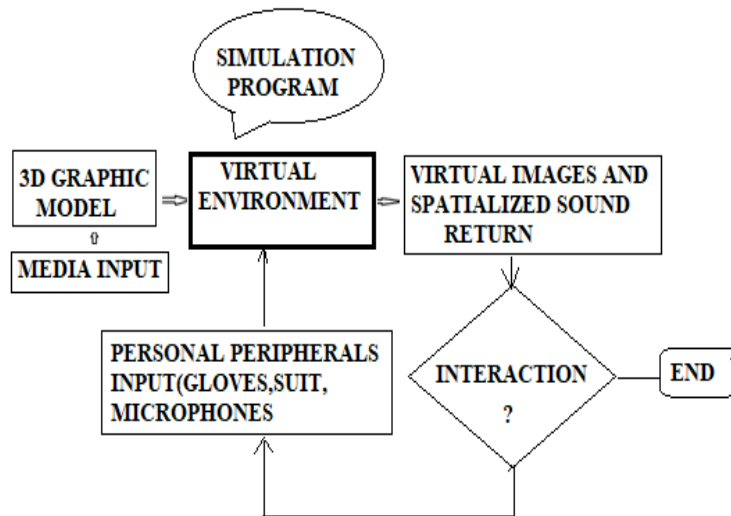
IV. VIRTUAL REALITY

Virtual reality technology refers to a computer-generated virtual world [30]. To immerse yourself in this multidimensional world, the user must wear a helmet containing a display that incorporates a sensor to track the movement and position of the wearer (HMD - Head Mounted Display). The sensors detect the position and movement data, which are used to update the image of the virtual world. In this way, the user is projected into this virtual world and can interact with the objects present in it, for example, simulating carrying out usual activities or driving a car or running. By manipulating three variables (space, time, and interaction) and the availability of a graphic interface, it is possible to create a dimension characterized by a strong sense of reality, whereby the subject believes he is in that world and can interact with it.

The virtual environment must have the following characteristics:

- Perception of really being in that world. The use of special equipment amplifies this feeling: software capable of reproducing 3D environments, a virtual reality viewer, integrated audio systems that offer surround support.
- The possibility of interacting with movements of the body, head, and limbs increases the feeling of taking possession of that dimension. For example, cyber-gloves, virtual limbs, joypads, etc., allow the user to touch, move, manipulate, or make virtual objects as if they were real.

The relationship between presence and immersion gives cognition in virtual reality. The first term refers to the level of psychological realism that a subject experiences from interacting with the virtual world. From the wise point of view, the second term means the ability of the environment to involve the senses of the subject, isolating it from the stimuli of the real environment [31].



Functional diagram of a VR system

The functionality of a virtual reality system requires different tools: modeling programs with which objects for the virtual world are created; moreover, to generate dynamics in real-time, simulation systems that follow natural laws are needed to move objects in the virtual world. For example, robust programs for viewing 3D images and spatialized sounds in real-time are also used. In addition, traditional systems used to communicate with a computer are no longer suitable for the multidimensional virtual world.

It was necessary to create new interfaces to interact with the computer. They are essentially based on two components:

- Sensors for controlling the virtual world;
- Effectors to provide feedback to the user.

These interface mechanisms are found in apparel, head-mounted displays (HMDs), and 3D sound systems.

Researchers have been developed methods to dynamically measure human movement in real-time through clothing equipped with sensors that track the full range of actions of specific activities carried out by the wearer. The movement is recorded, digitized, and then sent to the computer, which displays them on the screen, replicating all the actions performed in real-time.

This clothing consists of gloves and suits. Gloves are interactive devices, similar to a typical glove worn on the hand, facilitating tactile sensing and refined motion control. Tactile sensing involves simulating the sense of touch and includes the ability to sense pressure, linear force, temperature, and surface structure. Through the control of the fine movement achieved with sensors, the actions of the user's hands and fingers are detected. These movements are then translated into signals that can be used by a virtual hand to allow the user to interact normally with the objects in the virtual environment. The suit wraps the body with

directional sensors to transmit electrical signals corresponding to the wearer's movements to a computer. These signals are then converted into digital form, and the resulting data are displayed in the virtual environment. We often use gloves in conjunction with a position and orientation sensor that allows the computer to determine spatial coordinates. They can also be used in conjunction with suits and associated with HMD.

Finally, the viewers called HMD are devices that allow the wearer to immerse themselves in virtual reality. They include a head-mounted stereoscopic display and provide separate images for each eye, a stereo sound, and an ahead motion tracking sensor.

We can distinguish three types of virtual reality:

- Immersive Virtual Reality (RVI)), characterized by the intense sensation of immersion in virtual space thanks to the interaction with the virtual objects present.
- Non-Immersive Virtual Reality, the simulated environment does not stimulate the sense of immersion; it is perceived as the real one.
- Augmented Reality (AR) enriches the objects in the real world with perceptual information produced by the computer through multiple sensory modalities (visual, auditory, tactile).

Below is a brief overview of the state of the art

V. BACKGROUND

This section reports significant experiences developed based on neural networks combined with virtual reality. Through the analysis of facial expressions, it is possible to understand the emotional state. Neural networks coupled with virtual reality represent a valid platform for analyzing human behavior and effectively treating individuals with autistic disorders. Virtual reality (VR) offers distinct advantages over conventional

visualization approaches. Applications range from the entertainment industry to the rehabilitation field.

In this scenario, one of the first applications was that of Belič [32], who developed a methodology in which the two technologies are used to model complex polycrystalline materials. The neural network is used for generating the grain of the material. It makes the manipulation of the grain easier and the representation of the same very compact. It represents the preliminary phase for realizing the polycrystalline material model for virtual reality. The models perform various system optimization or control activities by simulating reality. To make a realistic model of the observed material, the shape of the grains represented by the neural network must come as close as possible to the observed material. The grains are first generated, then, based on the properties of the observed material, the grain shape optimization process is employed to get closer to the observed sample. With the VR approach, information is obtained that is not accessible with traditional techniques (characterization and analysis techniques). The VR must have the model accompanied by information on the mechanical and electrical properties and on the shape of the grains. The use of virtual reality made it possible to obtain additional information such as life expectancy, the diffusion process, or the cracking of the material. VR is used to predict any anomalies produced in the material; it allows you to view the discrepancies detected concerning the expected model. Finally, the virtual environment corrected the model to provide better results. The ultimate goal is to obtain the virtual roughness of the grain as close as possible to the desired one.

An interesting experience is developed by Nino et al. [33] on a VR mobile system based on Neural Network to an IMU gesture controller to simulate the feeling of embodiment [34] and presence. The application aims to demonstrate that combining a sound and motion controller with haptic feedback can improve immersion on mobile interfaces. The interface recognizes the user's gesture (vocal sound and movement) and displays it during execution. The visualization is managed through the gesture controller to improve presence and embodiment. To provide the user with a smooth display of the gesture produced, the authors have created a series of reference gestures that represent the best possible execution of that gesture. To make the set of these "ideal" gestures, a motion capture system was used to record the movement and voice produced by a shintaido expert while performing each gesture [35]. Each ideal gesture is played back for the user while performing the corresponding gesture. Through the dynamic control of the speed of reproduction of the gesture, the user features a degree of control over the avatar's movement. When the user has executed his movement, the captured gesture is classified, and the appropriate response is triggered.

Finally, the authors used a multilayer neural network (ANN) for speech and motion recognition. The network architecture and weights have been imported into the Unity3D platform. Without wishing to be a replacement solution for a dedicated tracking system, it can represent a low-cost method of controlling the gestural execution of a subject.

In the field of hand gesture recognition, there is an application developed by Fu and Yu [36] in which they use a neural network for the classification of gestures and, to evaluate the accuracy of recognition in real-time, they build a Unity scene using the data collected by the IMU. Hand gestures are an alternative to other natural methods of Communication. The possibility of recognizing the movement and the pose of the hands allows the observer to understand the intentions of who performs them. Hand gesture recognition is used in many applications such as sign language recognition, sign recognition for controlling robots, and augmented reality. The authors used the IMU as a data input source and structured the application into the following phases:

- Collecting training and testing data;
- Extracting features from IMU inputs;
- Building neural network-based classifier models;
- Building test visualization in Unity.

VR Arduino is used as the data input source. It contains an IMU that provides 9 degrees of freedom sensor readings, including gyroscopes, accelerometers, and magnetometers. It is connected to the PC via USB using Teensy.

The sensor data is recorded on the PC as a training data set and used directly in real-time tests.

The user can hold the Arduino chip in his hand and perform any hand gesture patterns.

A 3D LSTM convolution was used as a neural network to acquire better gestures and obtain high recognition accuracy.

To determine which features work best with the model, they experimented with several combinations of quaternion and raw sensor inputs. The best performing characteristics were the combination of quaternion, gyroscope, and accelerometer data. IMU test data from VRduino are transferred directly into Unity via serial port to visualize the classification.

The tests found that the classifier can predict the user's gesture with reasonable accuracy in real-time.

Sait and Raza [37] presented a prototype based on an innovative hybrid technology in which are present VR and ANN models to build a VE environment in which users can meet and chat together and at the same time feel like they are in a real environment. From a conceptual point of view, the model starts from real-world situations and then, through the IVR and ANN model, creates a virtual world in which the participants are the managers of the model, using the Internet as a

medium to connect the IVR model. The model components are thus: Immersible Virtual Reality, Artificial Neural Networks, Internet, and Database.

For the need to have a system design for a realistic visualization, the authors used the IVR. It is the technology model which allows people to immerse themselves in the artificial environment.

ANN is the architecture used to simulate the image and express the user's emotion in the VE. Users' voices, videos, and images are input to train the ANN. Those data are also used for security purposes to check the users' identities.

The Internet connects people spread over different points into a single point of contact on the web. The ANN network is trained on the Internet with a sample of users. The Internet allows scalability for a large number of users and ensures reliable transport of a large amount of multimedia data.

The database must contain a large number of quality video samples and possess a large number of facial recognition expressions to store the video and audio data of the participants.

The model requires two implementation steps. The first is for the interface, and the user inserts his data into the model, followed by the loading of their image, voice. These data will be stored in the audiovisual database to the interface to configure the user with the model. Later, when the user enters the system, the interface will check their identity with the stored data. Finally, the interface must be designed using 3D graphics software [38].

In the second phase, the IVR will provide the chosen environment to those users participating in the network, and the ANN will provide simulated images of the users.

As noted above, there is a growing need for non-experts to understand how deep learning works. For their complexity, neural networks are seen as black boxes. New techniques are experimented with to make the understanding of their functionality as accessible as possible.

In this context, Meissler et al. [39] have developed a technique for visualizing convolutional networks in virtual reality. Their solution is mainly aimed at new designers and aims to provide them with a basic understanding of the functionality of networks. The solutions currently available are essentially usable either by developers of deep learning systems or by those who already have detailed knowledge of deep learning processes or by those who generally have an interest in interactive visualization but are not attracted by the representation capabilities of virtual reality.

Their solution stems from using the advantages of virtual reality in visualization compared to traditional 2D or 3D tools available on a standard desktop screen. Their approach explores the graphic effects produced by the immersion feature. The visualization is intended

to illustrate the structure and functionality of the network. For example, the user within the Unity platform can select different inputs, modify the functional structure by inserting or removing a layer and obtain their classification by CNN.

The network model used is LeNet5 [40], whose structure is as follows:

- 2 Convolutional layers.
- 3 Fully connected layers.
- 2 Average pooling layers.
- Tanh is the activation function for the hidden layer.
- SoftMax is the activation function for the output layer.
- Cross-entropy as cost function.
- Gradient descent as optimizer.
- 60000 trainable parameters.

The model was defined in Keras [41] using PYTHON.

The convolutional and pooling layers were displayed as boxes whose dimensions were based on the layer's size. This type of representation should allow the user to see the structure of the model and the dimensional change of the data. Feature maps of each convolutional and pooling layer are rendered 2D images with matplotlib using the gray colormap.

The experiments carried out on the participants who did not know the topic confirmed the model's validity. Some of them declared that the application was completely self-explanatory. The experiments wanted to demonstrate that virtual reality compared to traditional means of learning, does not distract people but, on the contrary, allows them to focus more on the subject and even longer due to the involvement produced by the virtual environment.

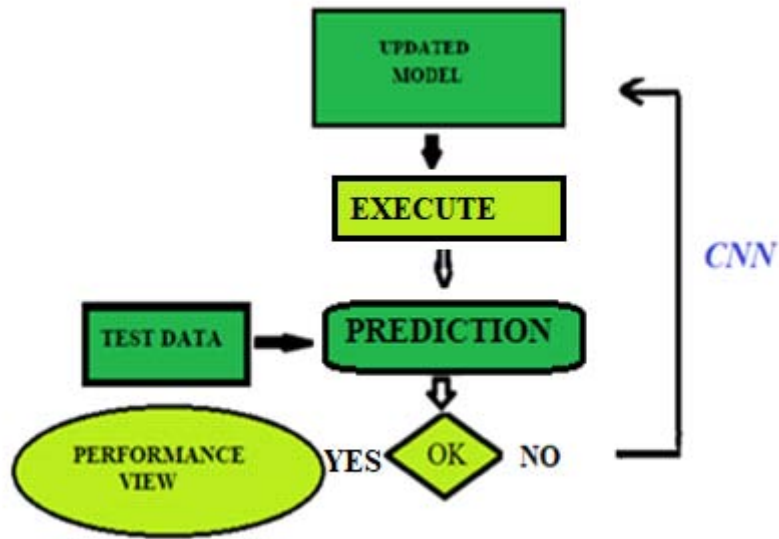
VI. METHODOLOGY

Our solution is based on a VR platform where users can develop deep convolutional neural network models for image classification. This type of approach allows verifying the components of a network and becomes a good training tool for professionals who want to design deep neural networks.

In our work, we want to show the opportunity to use the Unity platform (2019) to create a virtual reality environment. We believe it offers some advantages not found in various visualization systems in 2D and 3D desktop screens. Most of the existing platforms do not provide, for inexperienced users, simple interfaces through which it is possible to vary the network architecture and adjust the model parameters. For example, the VR platform allows the user to define the network architecture by specifying the sequence of the layers. The user can interactively verify its operation and decide based on the results obtained by modifying it. We have used this environment to verify the functioning and possibly improve the prediction level of a

convolution neural network to recognize human activities (HAR). In addition, the user can select input data and have them classified by CNN. In addition, we previously

created a network to identify human activities (HAR) with the STMicroelectronics AI STM32 module [42].



Inference model

a) CNN model

Convolutional Neural Networks (CNN) represents an artificial neural network architecture that has found wide application in identifying, with a certain probability, the content of the image provided by a computer. From a functional point of view, convolutional neural networks simulate how the human brain processes. Convolutional neural networks are built to analyze images included within certain data sets and classify objects in images within them. Each class of things must have a specific network previously trained for the reconstitution of that class. Finally, they process the data through multiple layers placed between the input and the output.

In our study, we have used a CNN to recognize the activities carried out by people daily, such as downstairs, upstairs, sitting, standing, walking. It is mainly used as a supervised learning algorithm and can

classify images from a sample dataset. It is based on the principle of the convolution filter applied to the pixel matrix that composes the image. Convolutional layers produce small transformations on images. Each convolution transforms a section of the image into a single value. The Kernel applied sequentially to the embodiment can highlight the fundamental characteristics of the image and, in this way, produce essential data for the classification of the image itself. Finally, we measure the model's performance through the accuracy and the loss.

A Keras model allows building the CNN network, created with the STMicroelectronics AI STM32 and trained on the public dataset Wireless Sensor Data Mining (WISDM). **The Keras DL model used is Shahnawax/HAR-CNN-Keras.** The code for the model definition is in the following listing:

CNN model

```

def cnnModel():
    model = Sequential()
    # adding the first convolutional layer with 32 filters and 5 by 5 kernal size, using the rectifier as the activation function
    model.add(Conv2D(numFilters,(kernalSize1,kernalSize1),input_shape=(numOfRows,numOfColumns,1),activation='relu'))
    # adding a maxpooling layer
    model.add(MaxPooling2D(pool_size=(poolingWindowSz,poolingWindowSz),padding='valid'))
    # adding a dropout layer for the regularization and avoiding over fitting
    model.add(Dropout(dropOutRatio))
    # flattening the output in order to apply the fully connected layer
    
```



```
model.add(Flatten())  
# adding first fully connected layer with 256 outputs  
model.add(Dense(numNeuronsFCL1, activation='relu'))  
#adding second fully connected layer 128 outputs  
model.add(Dense(numNueronsFCL2, activation='relu'))  
# adding softmax layer for the classification  
model.add(Dense(numClasses, activation='softmax'))
```

Fig. 1 shows the network architecture graph, while figures 2 and 3 show, for example, the information relating to the selected layers Conv2D and Dense obtained by clicking directly on the chart.

This information includes:

- the type of the layer;
- the name of the layer, whether the layer is trainable;
- what the data type is.

For Convolutional layers:

- the number of filters;
- the kernel size;
- the strides;
- padding;
- data format and dilation rate.

The activation function, whether bias is used, and how the kernels and (if applied) biases are initialized.

Each layer has its own unique set of characteristics.



Fig. 1: CNN architecture

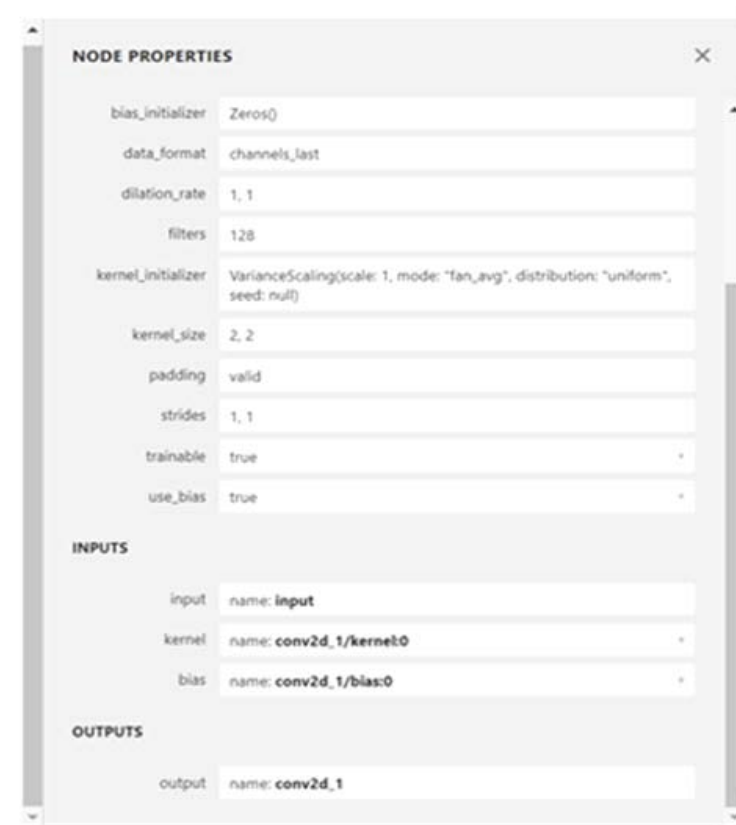


Fig. 2: Conv2D

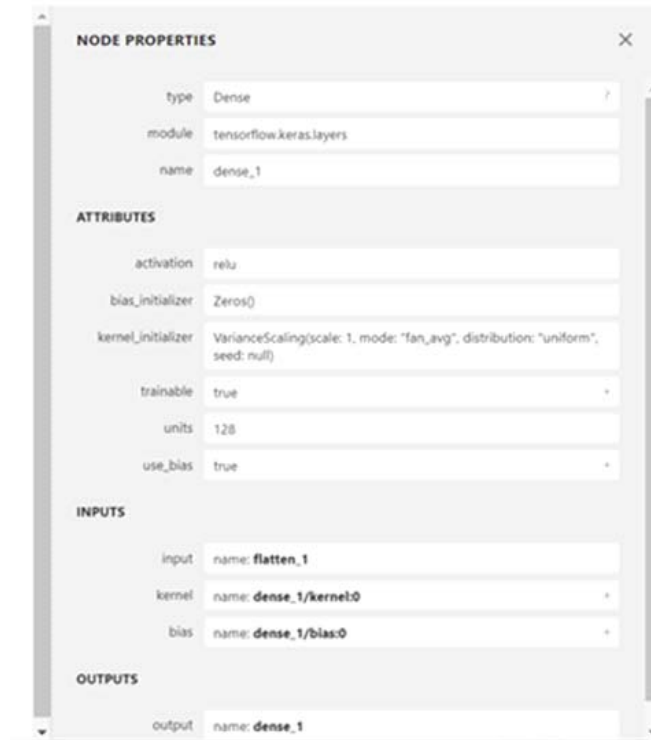


Fig. 3: Dense1

b) *Environment*

Within Unity, we used the Barracuda package developed by Unity Labs to develop our application. It is a cross-platform neural network inference engine. Pre-trained neural networks can be imported and run in Unity. In addition, it supports neural networks trained with PyTorch, TensorFlow, Keras, and Caffe.

We executed the following steps:

- Add the onnx file to the project. It behaves like a normal resource;
- Load the model from the asset;
- Create the inference engine (the worker).;
- Run the model and get the results.

To transfer the network, it is necessary to transform its format, as Barracuda is based on the open-source data exchange format ONNX.

```
# network
net =
# convert model to ONNX
onnx_model = keras2onnx.convert_keras(net, # keras model
    name="example", # the converted ONNX model internal name
    target_opset=9, # the ONNX version to export the model to
    channel_first_inputs=None)
onnx.save_model(onnx_model, "example.onnx")
```

The network model in ONNX format is imported into Unity as a temp file (Fig.4), whose structure will appear in the following graph (Fig.5).

ONNX, which stands for Open Neural Network Exchange, is an open standard that allows you to transfer machine learning models from different frameworks to ONNX. This interoperability enables you to move between various machine learning frameworks quickly. In addition, ONNX supports all popular machine learning frameworks, including Keras, TensorFlow, PyTorch, and XGBoost.

ONNX allows you to have a standard graphical representation for various frameworks. The ONNX graph represents the model through different computational nodes and can be visualized using appropriate representation tools.

To perform the conversion, we used a script. An extract of which is represented by the following code:

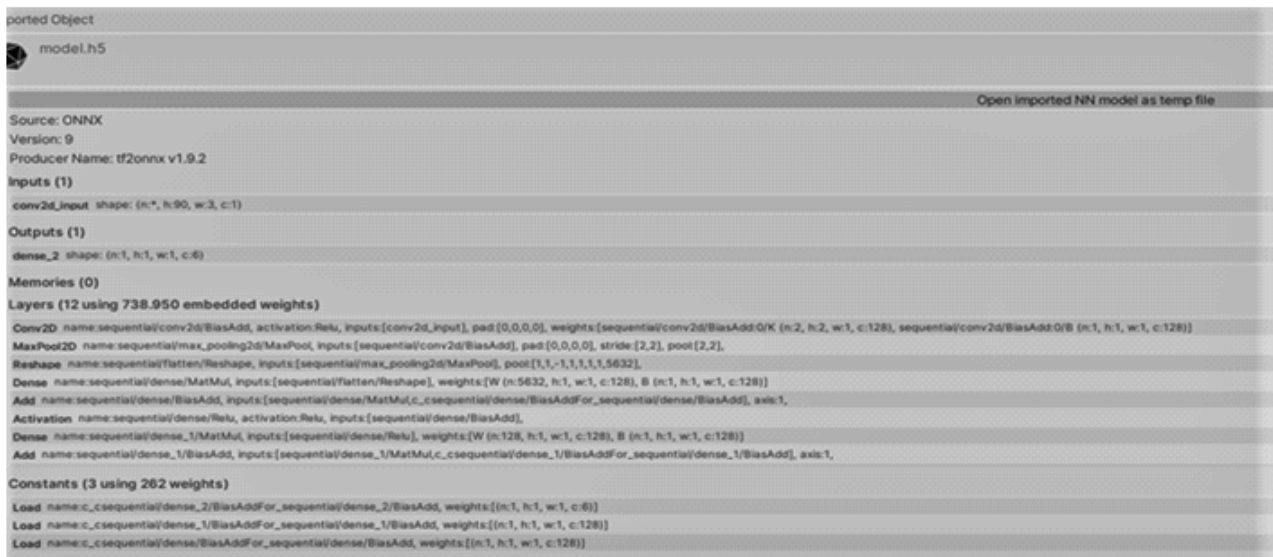


Fig. 4

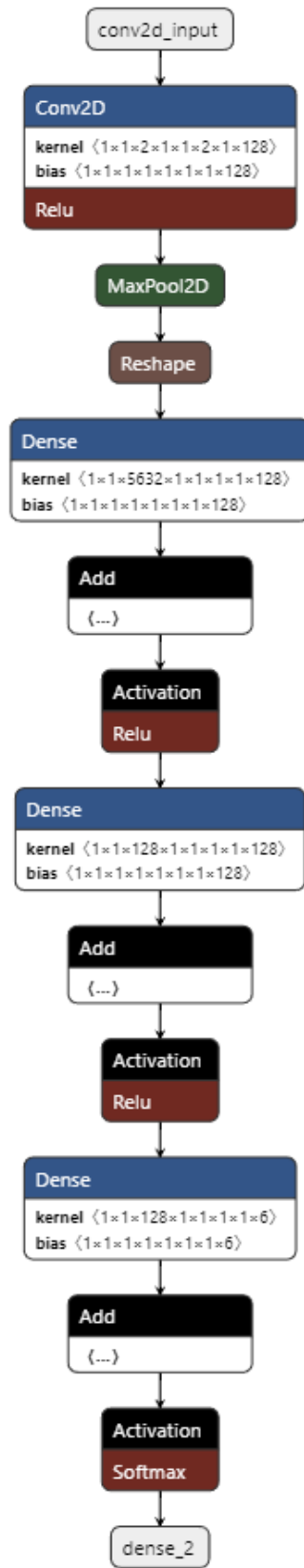


Fig.5: Barracuda model

As a next step, we need to create the inference engine. The core engine interface in Barracuda is called worker. IWorker breaks down the model into executable tasks and schedules them on GPU or CPU.

```
// GPU
var worker =
WorkerFactory.CreateWorker(WorkerFactory.Type.ComputePrecompiled,
NNmodel);
// CPU
Var worker= WorkerFactory. CreateWorker(WorkerFactory.Type.CSharpBurst,
NNmodel);
After loading the model and creating a worker, we move on to running the model:
Tensor input = new Tensor (batch, height, width, channels);
Worker. Execute(input);
After executing the model, we can collect the outputs through the instruction:
var output = worker.Peekoutput(outputName);
```

If the results are not satisfactory, we can repeat the execution of the model by varying the number of epochs or by modifying the structure.

In our case, to improve the accuracy of the network, we modified the structure by inserting an additional convolutional layer obtaining a value of 99%.

VII. DISCUSSION

Our approach shows how the VR technique can effectively design deep learning applications. Our platform represents a valuable tool for developers with little knowledge of machine learning, allowing rapid design. It is an interactive technique also valid as a learning tool. Compared to existing visualization techniques, Virtual Reality is a potential highly developed for neural network applications. Above all, it has proved to be suitable for classifying images in different scientific sectors. We can assume that it is applicable for any deep learning.

VIII. CONCLUSION

We have analyzed the most common visualization techniques used to represent neural networks, a valuable tool for their development. Still, we have found that some are sophisticated tools that are difficult to use for inexperienced designers.

Therefore, we have highlighted how to apply Virtual Reality to construct deep learning models. This type of representation, in fact, through the visualization of technical information on the network structure and on the data flow, represents an effective tool both for improving the forecasting capacity of neural networks and for training activities for those who do not have a profound knowledge of deep learning systems.

The scripts are different depending on whether you use a GPU or a CPU. We can create a Worker from the WorkerFactory. We must specify a backend and a loaded model.

We have experimented with applying it to a pre-trained neural network to recognize human activities (HAR). Still, we believe that it can effectively design all neural networks for image classification or object detection.

ACKNOWLEDGMENTS

This work was developed as an integral component of the project: "An integrated System for indoor people localization, tracking, and monitoring" supported by the Italian MIUR under GRANT PON Research and Innovation 2014-2020 Project Code C35E190000200001, AIM1839112-1: Technologies for living environments.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Shai Shalev-Shwartz, Shai Ben-David, "Understanding Machine Learning: From Theory to Algorithms," *Cambridge University Press*, 2014.
2. L. Arnold, S. Rebecchi, S. Chevallier, H. Paugam-Moisy, "An Introduction to Deep Learning," *European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, 2011.
3. P. G. Benardos, G. C. Vosniakos, "Optimizing feedforward artificial neural network architecture," *Engineering Applications of Artificial Intelligence*, 20(3), pp. 365-382, 2007.
4. I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning," *MIT Press*, p. 32, 2016.
5. G. Montavon, W. Samek, K. R. Muller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, 2017.
6. J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, "Understanding neural networks through deep visualization," *arXiv:1506.06579*, 2015.

7. D. Erhan, Y. Bengio, A. Courville, P. Vincent, "Visualizing higher-layer features of a deep network," *University of Montreal*, 1341, pp. 3, 2009.
8. M. Becker, J. Lippel, T. Zielke, "Gradient descent analysis: On visualizing the training of deep neural networks," *14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. pp. 338–345, 2019.
9. D. Liu, W. Cui, K. Jin, Y. Guo, H. Qu, "Deeptracker: Visualizing the training process of convolutional neural networks," *ACM Transactions on Intelligent Systems and Technology*, 10(1), PP. 1–25, 2018.
10. X. Chen, Q. Guan, X. Liang, L. T. Lo, S. Su, T. Estrada, J. Ahrens, "Tensorview: visualizing the training of convolutional neural network using paraview," *1st Workshop on Distributed Infrastructures for Deep Learning DIDL '17*, ACM Press, pp. 11–16, 2017.
11. M. Peters, L. Kempen, M. Nauta, C. Seifert, "Visualising the Training Process of Convolutional Neural Networks for Non-Experts," *CEUR-WS.org*, 2941, paper 108, 2019.
12. L. McInnes, J. Healy, J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv:1802.03426*, 2018.
13. M. Bock, A. Schreiber, "Visualization of neural networks in virtual reality using UNREAL ENGINE," *24th ACM Symposium on Virtual Reality Software and Technology, ser. VRST '18*, ACM, 2018, 132, pp.1–2, 2018.
14. N. Pezzotti, T. Höllt, J. Van Gemert, B. Lelieveldt, E. Eisemann, A. Vilanova, "Deepeyes: Progressive visual analytics for designing deep neural networks," *IEEE Trans. Visual. Comput. Graph.*, 24(1), pp.98–108, 2018.
15. M. Kahng, P. Y. Andrews, A. Kalro, D. H. P. Chau, "Activis: Visual exploration of industry-scale deep neural network models," *IEEE Trans. Visual. Comput. Graph.*, 24, pp: 88–97, 2017.
16. C. Olah. A. Mordvintsev, L. Schubert, "Feature Visualization," *Distill*, 2(11), 2017.
17. D. Erhan, Y. Bengio, A. Courville, P. Vincent, "Visualizing higher-layer features of a deep network," *Dept. IRO, Universite de Montreal, Tech.*, Rep, 4323, 2009.
18. A. Nguyen, J. Yosinski, J. Clune, "Multifaceted feature visualization: Uncovering the different types of features learned by each neuron in deep neural networks," *arXiv:1602.03616*, 2016.
19. A. Chattopadhyay, P. Manupriya, A. Sarkar, N. B. Vineeth, "Neural Network Attributions: A Causal Perspective," *arXiv:1902.023024*, 2019.
20. M. Kocaoglu, C. Snyder, A. G. Dimakis, S. Causalgan, Vishwanath, "Learning causal implicit generative models with adversarial training," *arXiv:1709.02023*, 2017.
21. S. Sattarzadeh, M. Sudhakar, A. Lem, S. Mehryar, N. Plataniotis, J. Jang, H. Kim, Y. Jeong, S. Lee, K. Bae, "Explaining Convolutional Neural Networks through Attribution-Based Input Sampling and Block-Wise Feature Aggregation," *Computer Vision and Pattern Recognition*, arXiv: 2010.00672.
22. S. Weber, D. Weibel, F. W. Mast, "How to Get There When You Are There Already? Defining Presence in Virtual Reality and the Importance of Perceived Realism", *Frontiers in Psychology*, 12, p. 1538, 2021.
23. J. Lee, M. Kim, J. Kim, "A Study on Immersion and VR Sickness in Walking Interaction for Immersive Virtual Reality Applications," *Symmetry*, 9(5), p. 78, 2017.
24. A. G. Abulrub, A. N. Attridge, M. A. Williams, "Virtual reality in engineering education: The future for creative learning." *2011 IEEE Global Engineering Education Conference (EDUCON)*, pp. 751–757, 2011.
25. Jeremy Sutton, "What Is Virtual Reality Therapy? The Future of Psychology", 2020.
26. J. Psołka, "Immersive training systems: Virtual reality and education and training," *Instructional Science*, 23 (5-6), pp. 405–431, 1995.
27. W. Rawat, Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, 29(9), 2017.
28. Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature*, 521, pp.436-444, 2015.
29. J. Li, J. Cheng, J. Shi, F. Huang, "Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement," *Advances in Computer Science and Information Engineering*, 169, pp 553–558, 2012.
30. C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, J. C. Hart, "The cave: audiovisual experience automatic virtual environment," *Communications of the ACM*, 35(6), pp. 64–69, 1992.
31. F. Biocca, B. Delaney, "Immersive virtual reality technology," *Communication in the age of virtual reality*, Lawrence Erlbaum Associates, Inc., pp. 57–124, 1995.
32. I. Beš, "Neural Network Modelling and Virtual Reality," *The Thousand Faces of Virtual Reality* cap.7, 2014.
33. Juan Nino, Jocelyne Kiss, Geoffrey Edwards, Ernesto Morales, Sherezada Ochoa, and Bruno Bernier, "Enhancing Mobile VR Immersion: A Multimodal System of Neural Networks Approach to An IMU Gesture Controller," *The Engineering Reality of Virtual Reality, IS&T International Symposium on Electronic Imaging*, pp. 184-1-184-5, 2019
34. B. Spanlang, J.-M. Normand, D. Borland, K. Kilteni, E. Giannopoulos, A. Pomes, M. Gonzalez-Franco, D. Perez-Marcos, J. Arroyo-Palacios, and X. N. Muncunill, "How to build an embodiment lab:

- achieving body representation illusions in virtual reality," *Frontiers in Robotics and AI*, 1, p. 9, 2014.
35. D. Franklin, "The Meaning of Tenshingoso," *Journal of the U.S. Shintaido Movement Issue*, 10, 2001.
 36. A. Fu, Y. Yu, "Real-time Gesture Pattern Classification with IMU Data," 2017.
 37. A. R. W. Sait, M. N. Raza, "A Virtual Environment Using Virtual Reality and Artificial Neural Network," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 2(12), n. 12, 2011
 38. Yang – Wai Chow, "A Cost-Effective 3D interaction Approach for Immersive Virtual Reality", *The International Journal of Recent Trends in Engineering*, 1(1), pp.529-531, 2009.
 39. N. Meissle, A. Wohlan, N.Hochgeschwender, A. Schreibe, "Using Visualization of Convolutional Neural Networks in Virtual Reality for Machine Learning Newcomers," *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2019.
 40. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11), pp. 2278-2324, 1998.
 41. F. Chollet and al., "Keras," <https://keras.io>, 2015.
 42. L. Bibbò, R. Carotenuto, "An integrated system for indoor people localization, tracking, and monitoring" *Journal of International Scientific Publications, Materials, Methods & Technologies*, ISSN 1314-7269, Volume 15, 2021





This page is intentionally left blank



Vehicle Routing Problem with Time Window Constraint using KMeans Clustering to Obtain the Closest Customer

By Jai Keerthy Chowlur Revanna & Nushwan Yousif B.Al-Nakash, PhD

Information Systems Engineering and Management & Harrisburg University of Science and Technology

Abstract- In this paper, the problem statement is solving the Vehicle Routing Problem with Time Window constraint using the Ant Colony Algorithm with K-Means Clustering. In this problem, the vehicles must start at a common depot, pickup from various warehouses, deliver to the respective nodes within the time window defined by the customer and return back to the depot.

The objectives defined are to reduce number of vehicles employed, the total logistics cost and to reduce carbon emissions. The mathematical model described in this paper has considered multiple pickup and multiple delivery points. The proposed solution of this paper aims to provide better and more efficient solution while minimizing areas of conflict so as to provide the best output on a large scale.

GJCST-D Classification: F.1.1



VEHICLEROUTINGPROBLEMWITHTIMEWINDOWCONSTRAINTUSINGKMEANSCUSTERINGTOOBTAINTHECLOSESTCUSTOMER

Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

Vehicle Routing Problem with Time Window Constrain using KMeans Clustering to Obtain the Closest Customer

Jai Keerthy Chowlur Revanna*, Nushwan Yousif B.Al-Nakash[†], PhD

*Information Systems Engineering and Management
Harrisburg University of Science and Technology

Email: jchowlur@my.harrisburgu.edu, nal-nakash@harrisburgu.edu

Abstract- In this paper, the problem statement is solving the Vehicle Routing Problem with Time Window constraint using the Ant Colony Algorithm with K-Means Clustering. In this problem, the vehicles must start at a common depot, pickup from various warehouses, deliver to the respective nodes within the time window defined by the customer and return back to the depot.

The objectives defined are to reduce number of vehicles employed, the total logistics cost and to reduce carbon emissions. The mathematical model described in this paper has considered multiple pickup and multiple delivery points. The proposed solution of this paper aims to provide better and more efficient solution while minimizing areas of conflict so as to provide the best output on a large scale.

I. INTRODUCTION

Transportation is one of the primary requisites of civilization and this fact continues to be true even today. In today's world of quick and safe deliveries, there has been a need for better service, reduction of vehicles used, maximizing profit, reduction in travel time variation and reduction of overall travel cost. To define these problems together, the term Vehicle Routing Problems was coined. This problem deals with the supply chain of an organization. Transportation is the backbone of the logistics of any organization and it takes up about 40 to 50 % of the total logistics cost, as stated in <https://www.cogoport.com/blogs/transport-cost> (accessed on 11 October, 2021). This includes international and domestic transport, customs, all modes of transport such as air, water, land and so on. It can be inferred that transportation cost is a major and important factor in the supply chain of an organization, so its cost optimization becomes a necessity. The logistics branch of the organization must work on the management of transportation, deliver within customer provided time frames, competing with other organizations for better service and service rates effectively, handling unpredictable events and so on.

The world is witnessing the digital growth spurt and along with its influence on almost every sphere of life and nature. Integration of logistics and e-business will be a fruitful endeavour. This incorporation will lead to

improvement in customer service, tracking, deliverance, time effectiveness as well as reduction in the overall cost.

Looking at the technical aspects of the Vehicle Routing Problem (VRP), there are initially p vehicles located at a depot that must deliver different amounts of supplies to q customers. Now, the VRP will aim to find the optimal route that a group of vehicles serve a group of users. This way a standard solution is obtained which contains all the routes that start and end at the depot, with the constraint that the goods are delivered within or before the time range set by the customer, capacity limit and the working time of the drivers are also considered.

This paper will discuss how the Ant Colony Optimization with K-Means Clustering (ACO-K-Means) has been employed to minimize costs when delivering goods from depot to the customer within or before the time frame constraint set. The mathematical model defined in this paper will tackle and solve the problems related to distribution, e-logistics, retail networks and so on.

Dantzig and Ramser [1] were the first ones to introduce the Vehicle Routing Problem in 1959. Their solution was based on Linear programming. It was a truck dispatching problem that dealt with the delivery of gasoline at gas stations. Later, [2] Clarke and Wright came up with the savings method and it was termed as the Clarke-Wright algorithm. Their practical methodology gave better results than the Ramser-Dantzig algorithm. This was because the latter algorithm simply linked the customer pairs that were close to each other, which means that only distance constraint was considered, while the former not only considered the distance constraint, but they also reduced the distance rather than linking the two customers to different routes. Fast forward to 1992, Daskin and Malandraki came up with the time dependent vehicle routing problem (TDVRP) [3]. Then another solution was introduced by Ichoua et al. [4] which had a step function along with a piecewise linear function of time distribution which was fulfilling the FIFO (first in first out) principle, which was defined by Ahn et al. with this, several research and studies

popped up. Some were utilizing route construction savings method and an insertion method to solve incapacitated TDVRP (time dependent with/without time windows), some had heuristic solutions [8-10], some metaheuristic algorithms [5-7] and others hyper

heuristics [11].Figure 1. shows a generalized view of how a VRP is solved.

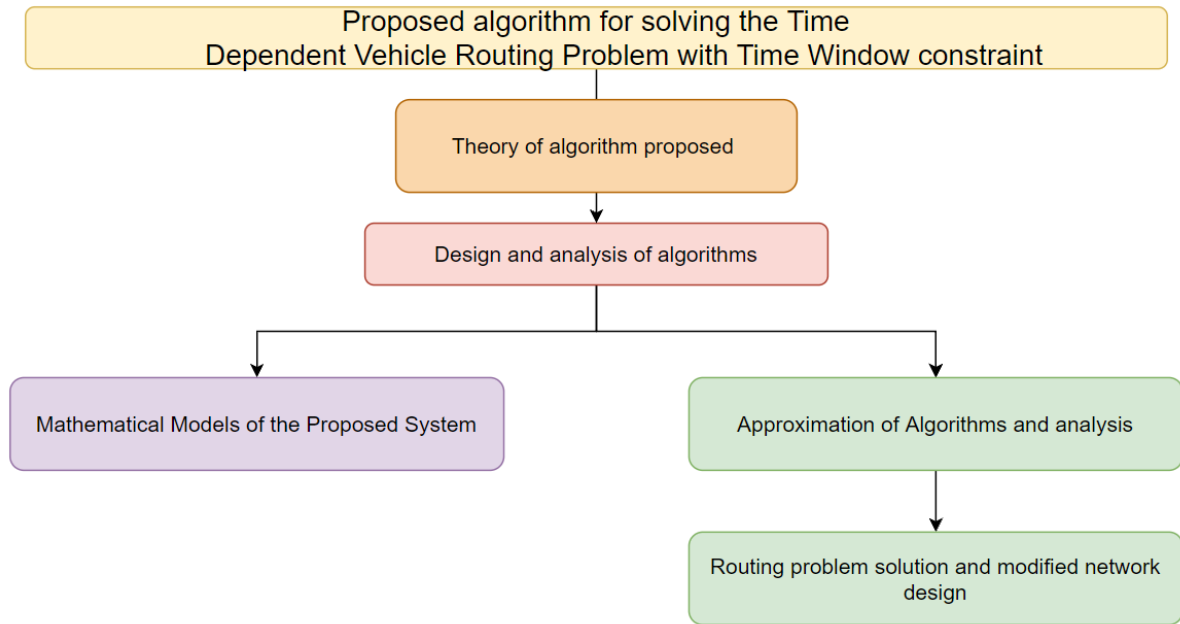


Figure 1: General VRP Solving Method

Many works of solving the VRP with the Time Window Constraint were inherited from the travelling salesman problem. The method used by the salesman to find the best and optimal route to deliver the goods to the respective customers from one or more depot and also take the goods from the customer back to the respective depots within the constraints set, has been extensively used in VRP, with the inclusion of extra constraints. Similar VRP variants have been mentioned below:

- Vehicle Routing Problem with the Time Window Constraint [12] that has been set by customers,
- Another modified VRP with the added constraint of using limited number of vehicles of varying holding capacity has been published as Mixed Fleet Vehicle Routing (MFVRP) [13],
- Another paper which has VRP with an added constraint where customers can request for delivery or pickup with the requirement that in every single delivery route, all pickups and deliveries to the customers are completed. This is known as Vehicle Routing Problem with Backhauls (VRPB) [13].

This paper has five sections in total. Section 1 deals with the introduction while section 2 deals with the literature survey. Section 3 handles the mathematical model of the proposed system [ACO using K-Means Clustering Algorithm], section 4 will explain the

approach to the solution, section 5 will have the results and case studies, with section 6 concluding the paper.

II. LITERATURE SURVEY

One of the heuristic solutions mentioned was provided by Hideki Hashimoto, MutsunoriYagiura and Toshihidelbaraki [8]. In their paper they generalized VRPTW by making travelling costs and duration to be time-dependent functions. They used local search algorithm to find the routes of vehicles and using that, evaluated a neighbourhood solution. they proposed an algorithm that could efficiently pick optimal routes using data from previous dynamic programming recursion that were used to evaluate the present solution. they even included a filtering method that determines which spaces in the neighbourhood are not to be searched so as to avoid dead ends in improving the solution. they finally conclude with a local search algorithm that combines all their modifications.

A metaheuristic solution was proposed by Yiyokuo[6]. In the research paper, the author has considered fuel consumption and carbon emission as the constraints to the Time-Dependent Vehicle Routing Problem (TDVRP). The paper has proposed an algorithm that determines a route that consumes less fuel and has the least carbon emissions. With this algorithm the author was able to provide an overall improvement of 22.69% in minimizing transportation

distances and 24.61% improvement in fuel consumption.

[11] has used a two-phase method that includes Genetic Algorithms along with Random Search incorporating simulated annealing concepts to solve the time dependent vehicle routing problem (TDVRP). This is a hyper heuristic solution.

Paper [14] has taken into consideration the problems of carbon pollution and environmental issues. Electric vehicles were considered to reduce the various problems mentioned but it brought along with it the issue of charging locations and battery capacity. To tackle these problems, a new variant in the classical VRPTW was brought about which integrated the ideas of multiple charging points that also have the facility of swapping batteries. The authors proposed a mixed integer programming model to tackle the issue using the improved ant colony optimization (ACO) algorithm with hybridised insertion heuristics and enhanced local search.

Another reference has been taken from [15] which is quite close in similarity with this paper's solution. The problem that the paper addressed was that deliverance of perishable goods within a given time frame was a daunting task and if unexpected events took place, the extremely important goods would not reach their destination, leading to a molehill of problems and difficulties. The authors Yao Wu, Bin Zheng and Xueliang Zhou have proposed a working model where the idea of disruption management has been employed to create a disruption recovery model with a different type split delivery that is used for inter-route recourse based on a previous TDVRPTW. It takes into account the nature of perishable goods and dynamic travel route choice in urban road networks. The, a tabu search algorithm is brought up to create a solution for the initial routing problem. This will be further extended to create the disruption recovery plan.

[16] Researchers have also used a novel ant colony optimization algorithm based on improved brainstorm optimization (IBSO-ACO) to solve VRP with soft time windows. According to this paper, the classical ant colony algorithm has been modified to efficiently solve the local optimum problem. Their research has given proof that it can achieve a lower routing cost at a high convergence rate than the classical ant colony (ACO) and the stimulated annealing ant colony algorithms.

Looking into other heuristic strategies involved, [17] has the space-filling curve with optimal partitioning

$$\sum_{o \in C_2} p_{om}^k - \sum_{y \in C_2} p_{yn}^k = 0 \quad \forall k \in K, \forall (m, n) \in R \quad (3)$$

A vehicle must pass starting and ending depots at least once and this is shown by equations 4, and 5.

as a solution while another has three-phase heuristics which has been developed by grouping a heuristic-based clustering algorithm solving VRP [18]. Summary of other important state-of-art modern heuristics is available in [19,20].

In this paper, we will be solving the Vehicle Routing Problem with Time Windows constraint using the modified Ant Colony Optimization with K-Means Clustering. Ants use pheromones to leave behind a trail for its comrades so as to use the optimal path fixed to reach the food source. There has been several research based on this behaviour of ants, such as [21], which was the first paper to be published on this topic. Papers [22-27] have various hybrid versions of ACO in varied fields.

Using this behaviour of ants and with the help of previous research work based on a somewhat similar problem, this paper aims to solve VRPTW using the K-Means Clustering algorithm to find the most optimal path to the customer.

III. MATHEMATICAL MODEL OF PROPOSED SYSTEM

This part will use certain terms and elements from [28]. It is a case study based on VRPTW regarding fresh food distribution centres. There will be two subsets of service nodes: pickup set P_S and delivery set D_S . The values of these terms are $|P_S| = x$ and $|D_S| = y$ respectively. Now, starting depot node is set to 0 and end depot is set to $(x + y + 1)$. A node will be replicated if it needs both delivery and pickup. Each vehicle has its set capacity and operation cost. If there is an order between pickup node m and delivery node, n then there will be a set R which contains pairs of (m, n) .

Looking at the objective function that minimizes total travelling cost, equation 1 is as follows

$$\min \sum_{k \in K} \sum_{m, n \in C} p_{mn}^k r_{mn} \mu_k \quad (1)$$

Here, K refers to the number of clusters and C refers to the centroid of clusters.

Equation 2 makes sure that each node is served by at least one vehicle

$$\sum_{k \in K} \sum_{n \in C_1} p_{mn}^k \geq 1 \quad \forall m \in C \setminus \{0, x + y + 1\} \quad (2)$$

Equation 3 showcases the constraint where the same vehicle k must pick and order from node m and deliver it to node n .

$$\sum_{n \in C_1} p_{0n}^k \leq 1, \quad \forall k \in K \quad (4)$$

$$\sum_{n \in C_2} p_{n, x+y+1}^k \leq 1, \quad \forall k \in K \quad (5)$$

If a vehicle reaches a node, it must leave it as well. This is shown in equation 6.

$$\sum_{m \in C_2} p_{mn}^k = \sum_{o \in C_2} p_{no}^k \quad \forall n \in C \setminus \{0, x + y + 1\}, k \in K, o \neq m, n \neq m \quad (6)$$

Equations 7, and 8 have integrated time constraints, subtour elimination and load constraints.

$$Tc_m^k + (t_{mn} + s_m) - \text{Big}M(1 - p_{mn}^k) \leq Tc_n^k, \quad \forall k \in K, n \in C_1, m \in C_2 \quad (7)$$

$$Lc_m^k + d_n - \text{Big}M(1 - p_{mn}^k) \leq Lc_n^k, \quad \forall k \in K, n \in C_1, m \in C_2, m \neq n \quad (8)$$

Now, if there is an order placed between two nodes and the pickup node must be visited before the delivery node, then equation 9 shows it.

$$Tc_n^k \geq R_{mn}(s_m + t_{mn}) + Tc_m^k, \quad \forall k \in K, m \in P_S, n \in D_S \quad (9)$$

Equation 10 shows time constraint while (11) shows capacity bound constraint.

$$l_m \geq Tc_m^k \geq f_m, \quad \forall m \in C \setminus \{0, x + y + 1\}, k \in K \quad (10)$$

$$(A_k + d_m, A_k) \geq Lc_m^k \geq (0, d_m), \quad \forall k \in K, m \in C \quad (11)$$

Now showcasing the constraint of limiting number of vehicles used and maximum working duration in equations 12, and 13.

$$\sum_{k \in K} \sum_{n \in C} p_{0n}^k \leq M \quad (12)$$

$$t_{max} \geq Tc_{x+y+1}^k - Tc_0^k, \quad \forall k \in K \quad (13)$$

This mathematical model is a small-scale solution. Because it being a population based metaheuristic solution, it is to be expanded step by step. Once the pheromone value is successful in small scale, the forging behaviour of ants could be applied to VRP in large scale for large number of vehicles.

IV. APPROACH TO THE SOLUTION

In this paper, the Vehicle Routing Problem with Time Window constraint has been resolved using a modified version of the Ant Colony Optimization using K-Means Clustering. Marco Dorigo was the first person to introduce Ant Colony Optimization, in the 90s, in his Ph.D. thesis. The solution algorithm is based on the behaviour of ants, the way they live in colonies and search for food. While an ant goes around, searching for food, it leaves behind pheromones that act as a beacon. It acts as a communication mechanism and each time the ant leaves a pheromone trail, it tells the other ants about the quality and quantity of food the former ant had been carrying. This way, there are several set paths that the ants use based on the number of pheromones released in a path. The shortest and fastest route is chosen for maximum traffic. .

Figure 2 shows process flow of the proposed solution. The purpose of the optimization model was to reduce specific energy consumption by taking into account the diameter of the pipe, the number of

sprinklers, the working pressure of the end sprinklers along the pipeline, and the coordination conditions of the pump pipeline. A pressure drop between adjacent sprinklers has been added to the heuristic function of ACO's design to reflect the distance between the two cities of the Traveling Salesman Problem (TSP).

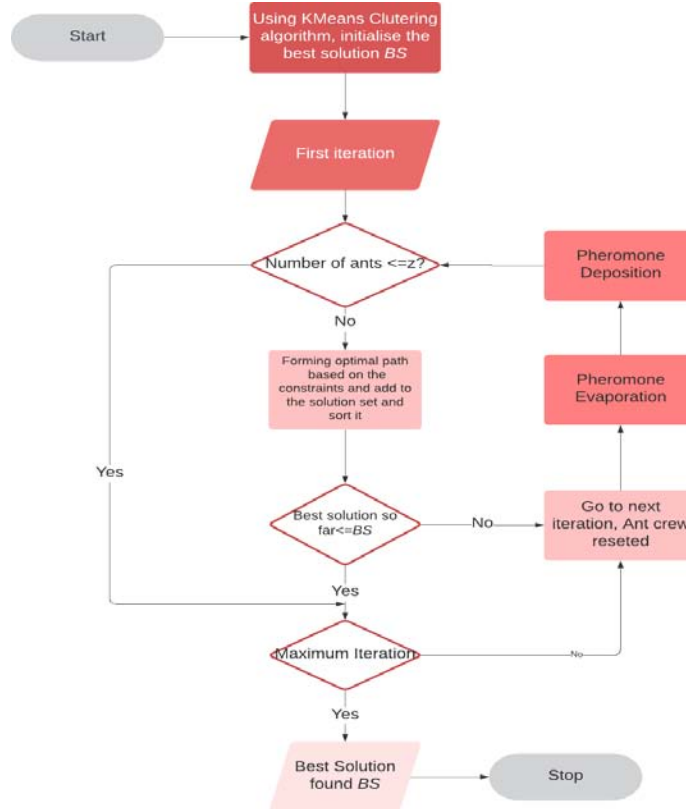


Figure 2: Process Flow of the Proposed Solution [Reference number]

As shown in the above figure 2, Ant Colony Optimization (ACO) algorithm is a probabilistic technique based on the above phenomenon to find the optimal path. With the inclusion of K-Means Clustering, this modified approach has solved the constraints of the MPMDVRPTWHF, which has resulted in shorter time consumption, delivery within the time window and lower transportation costs along with the inclusion of multiple pickup and delivery nodes wherein a pickup point might or might not have multiple delivery locations.

a) *Parameter initialization*

Looking at the research done in [28], the following parameters are set as follows:

Number of ants $z = 22$, $\alpha = 2$, $\beta = 5$, $\rho = 0.80$, $\theta = 80$, *elitist ants* $\sigma = 3$.

In the graph $G = (N, A)$, each arc (m, n) has been assigned a variable called pheromone trail τ_{mn} . The probability of better solution is directly proportional to the pheromone intensity. This means that when an ant wants to go to another node from its current node, it will choose the one with the maximum pheromone intensity. to make this work, a fixed quantity of

pheromone is allocated to every arc. To decide which node to proceed to (node n), the k th ant will use the pheromone trail τ_{mn} which is showcased below in equation 14:

Initially, all probabilities are set to 1. $\lambda = \frac{1}{\tau_{mn}}$ is a heuristic value, pheromone concentration on the edge when the ant travels from node m to node n is denoted by τ_{mn} and relative influence of the pheromone concentration and the heuristic value is shown by α and β .

If we go into the specifics, then λ_{mn} denotes how much favourable is the next node n while τ_{mn} implies how much better is the next node relatively.

b) *Solution Construction*

In this scenario, the solution is generated when an artificial ant takes vehicles from the vehicle set and constructs a path, starting from the warehouse or depot, by choosing those nodes that satisfy the set of constraints. The ant continues to build the route until the limit of route length has been reached or when the time window constraint has been disobeyed. so, in forming the route, the ant will check each node whether it fulfils all the constraints and if it finds such a node, it will

append it to the route, update the variables and go for the next node, using the updated variables. These changes in each iteration are all recorded in a solution set, which will then be used for finding the best solution from the set.

When determining the optimal route for the best solution, pheromone update is used which includes pheromone deposition and pheromone evaporation. Pheromone update is used to elevate the pheromone values that are found on good solution paths and

Where trail persistence $1 \geq \rho \geq 0$ of the evaporation factor $1 \geq \left(\frac{\theta}{S_{avg}} + \rho\right) \geq 0$ and θ is a constant. In each iteration, z number of ants find

decrease those that are on bad solution paths. In pheromone deposition and evaporation, pheromone values either increase or decrease at a constant rate [29].

The pheromone evaporation equation is given as such,

$$\tau_{mn} = \tau_{mn} \left(\frac{\theta}{S_{avg}} + \rho \right), \quad \forall (m, n) \in A \quad (15)$$

$S_{avg} = \sum_{k=1}^z \frac{S^k}{z}$ of average total distance. Then pheromone is updated by the elitist and best ants

After the evaporation process, only the best and the elitist ants can update the pheromone deposits on the optimal path chosen, which is given by the equation

$$\tau_{mn} = \tau_{mn} + \Delta \tau_{mn}^* + \sum_{b=1}^{\sigma-1} \tau_{mn}^b \quad (16)$$

Where,(16)

$$\Delta \tau_{mn}^\lambda = \begin{cases} \frac{\sigma - \lambda}{S^\lambda} & \text{if } \lambda \text{th best ant travels on edge } (m, n) \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$$\Delta \tau_{mn}^* = \begin{cases} \frac{\sigma}{BS} & \text{if best part of solution is } (m, n) \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

Looking at equations 17 and 18, it can be concluded that there are two types of pheromone depositions that are deposited on the trails during the pheromone update process. First, if σ elitist ants have travelled a path, that path will be updated as the best solution so far (*BS*), in accordance with the ACO+K-Means Clustering algorithm. $\Delta \tau_{mn}^*$ denotes the pheromone update by the elitist ants. Second, out of the z ants available, only $(\sigma - 1)$ best ants, in the current iteration, can deposit pheromone on the path they have traversed. The term $\Delta \tau_{mn}^\lambda$ is used to denote the pheromone quantity laid down on the trails that have been traversed by them and the amount of pheromone that have been deposited by the ants are determined by their solution quality S^λ and rank λ and the value is equal to $\Delta \tau_{mn}^\lambda$. To summarize, the elitist ants need to increase the probability of the best-solution so far after each iteration as the values that are updated will act as reference values for the next iteration. The ranking methodology has been employed in [17] so as to reduce pheromone deposition on those routes that have relatively lesser favourability.

ready time. This section will be in comparison with [30] as it has used the same data set. This comparison will help in proving that the proposed solution from this paper is the better method of solving the (MPMDVRPTWHF) as it gives better cost reduction with lesser percentage of carbon emissions, along with optimized fuel consumptions and lesser vehicles used.

b) Parameters Defined

The parameters defined in this paper are derived from [30] as this paper is in comparison with the latter. Similar to [30] the delivery vehicle used is a refrigerator car and the set of pre-defined parameters as shown in table 1 given below.

V. CASE STUDY

a) Dataset Used

The dataset for this paper has been taken from the Solomon-100 standard test set which have 20 problem cases. It also includes x-y location coordinates, service time, demand by customers, due dates and

Table 1: Parameters to Be Used Throughout this Paper

| Parameters | Implications | Predefined Value |
|-----------------|---|-------------------------|
| g | Gravitational constant of Earth | 9.81 kg/m ² |
| P | Density of atmosphere | 1.225 kg/m ³ |
| θ | Gradient | 0 |
| ΔT | Temperature Difference between container of vehicle and atmosphere | 20° C |
| λ | Frequency of opening the vehicle door | 0.6 |
| t_c | Transportation price of per unit weight of goods | 1 |
| δ | Decline coefficient of product freshness | 0.01 |
| ϵ_1 | Early arrival time penalty coefficient | 0.6 |
| ϵ_2 | Late arrival time penalty coefficient | 0.8 |
| R_n | Number of iterations of pheromone renewal | 20 |
| α | Weight of pheromone concentration | 1 |
| β | Weight of heuristic function | 3 |
| ρ | Volatility coefficient | 0.8 |
| q^f | The parameter of the selection rule of Pseudo-random proportional action | 0.9 |
| τ_{min} | Minimum value of pheromone | 0.001 |
| τ_{max} | Maximum value of pheromone | 10 |
| w_1, w_2, w_3 | Weights of optimization objectives | $\frac{1}{3}$ |
| R | Fuel consumption per unit time ($\frac{g}{s}$) | 22.37 |
| h | Fuel consumption and carbon emission conversion factor ($\frac{L}{kg}$) | 1/2.7 |
| δ | Fuel-air ratio | 1 |
| φ | Engine friction factor | 0.2 |
| T | Engine speed of refrigerated truck(r/min) | 2000 |
| V_s | Air Displacement(L) | 1.051 |
| η | Effective power transmission | 0.45 |
| μ | Energy consumption constant | 44 |
| M_q | Refrigerated truck weight(kg) | 1510 |
| A | Windward area of vehicle(m ²) | 3.96 |
| C_d | Coefficient traction | 0.4 |
| C_r | Coefficient of rolling resistance | 0.01 |
| γ | Heat transfer coefficient refrigerated truck | 0.4 |
| S_w | External surface area of refrigerated truck(m ²) | 43.36 |
| S_v | Interior surface area of refrigerated truck(m ²) | 22.32 |
| | Vehicle maximum load | 200 |
| | Vehicle fixed cost | 200 |
| | Soft time window | 20 |
| | Carbon tax | 20 |

c) Result Analysis

The entire result section has used the Pareto optimal principle for obtaining the solution. The Pareto Principle states that 80 percent of a project's benefit comes from 20 percent of the work. The optimal version of it makes the sub objectives suppressed so as to efficiently solve the main objective. Due to this there is very little scope of conflict of objectives from the sub objectives and a noiseless solution id obtained. Referring to [30], this paper the objectives chosen will be carbon emission reduction, total cost, time frame and customer satisfaction.

Using several test cases of 25,50,75 and 100 customers in three different scenarios, the proposed ACO algorithm with K-Means clustering provides a better solution in comparison. The results are arranged in the Pareto optimal solution format. The test cases and their outputs are given below from figure 3 to figure 26.

Table 2: Shows the Legends used in this Paper

| | |
|-----|-------------------------|
| CNY | Cost in CNY |
| CE | Carbon Emissions |
| CS | Customer Satisfaction |
| NY | Number of Vehicles Used |

Table 2 shows the used legends as explained in figure 3, figure 5, figure 7 and figure 9.

i. 25 Customers

C101(25) test table is used here for obtaining the most optimal path with better results of the constraints set. This solution has used the Pareto optimal approach and figure 3 has shown the comparison between [3] and this paper. It is clearly

visible from the graph that the proposed algorithm of ACO+K-Means [PS_KPSO] clustering has better output in terms of carbon emission, customer satisfaction and total transportation cost. ACO (Ant Colony optimization) and ACOMO (Ant Colony optimization Multi objective) which is heuristic function.

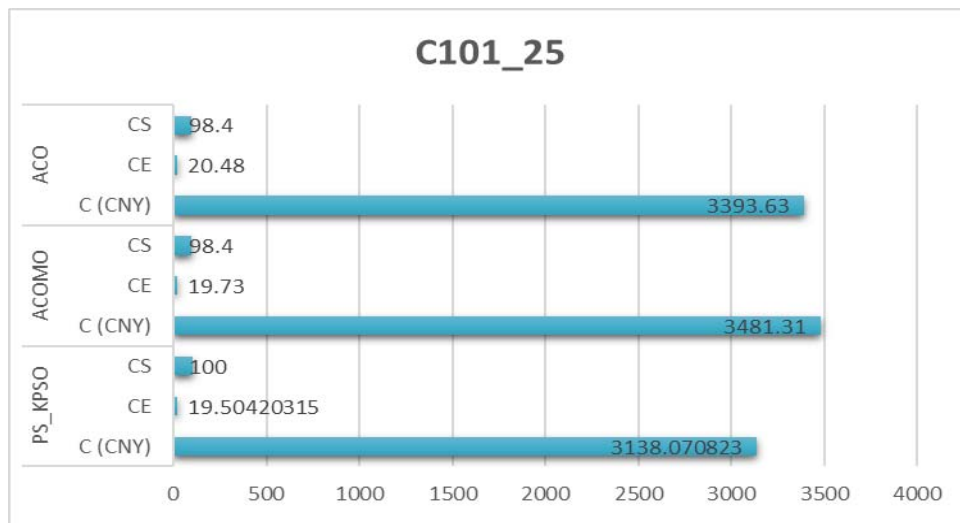


Figure 3: c101(25) comparison with [30] (ACOMO) and ACO+K-Means Clustering (PS_KPSO)

PS_KPSO algorithm has given total cost as 3138.07, lower carbon emission of 19.50 and 100 percent customer satisfaction in all cases. Below in figure 4, showst he trucks used is 3 with respective optimal travel paths (0, 5, 3, 7, 8, 10, 11, 9, 6, 4, 2, 1,0), (0, 13, 17, 18, 19, 15, 16, 14, 12,0) and [(0, 20, 24, 25, 23, 22, 21, 0)

This part has used the c101(50) test table. 5 trucks have been employed with respective paths (0, 43, 42, 41, 40, 44, 46, 45, 48, 50, 49, 47,0), (0, 5, 3, 7, 8, 10, 11, 9, 6, 4, 2, 1,0), (0, 20, 24, 25, 27, 29, 30, 28, 26, 23, 22, 21,0), (0, 32, 33, 31, 35, 37, 38, 39, 36, 34,0) and (0, 13, 17, 18, 19, 15, 16, 14, 12, 0). The end results are 33.43 for carbon emissions, 5942.72 cost and 100 percent customer satisfaction. Figure 5 shows the comparison between [30] and this paper results while figure 6 displays the routes taken by the 5 trucks.

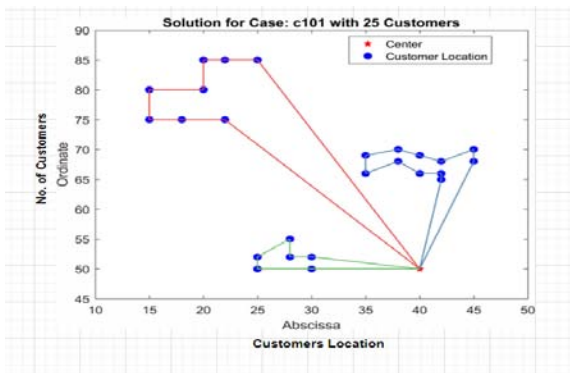


Figure 4: c101(25) vehicle distribution route50 Customers

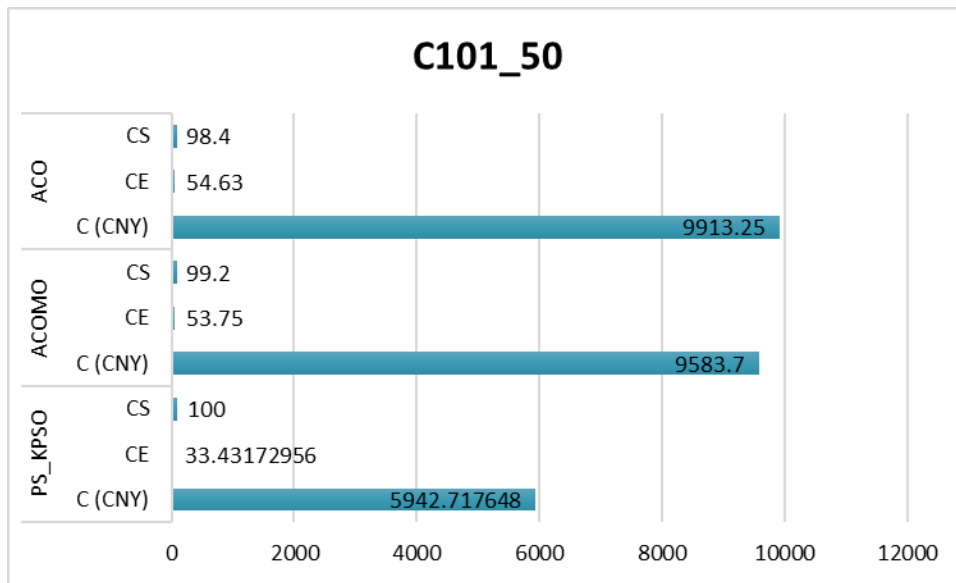


Figure 5: c101(50) comparison with [30] (ACOMO) and ACO+K-Means Clustering (PS_KPSO)

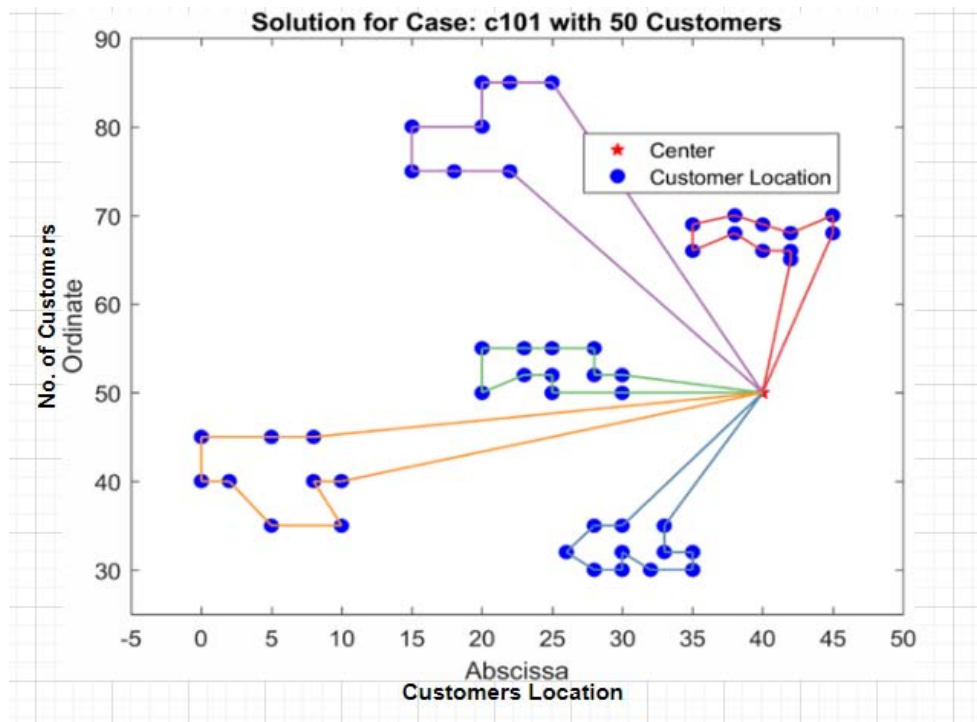


Figure 6: c101(50) vehicle distribution route

ii. 75 Customers

The c101(75) dataset has been used in this part. The number of vehicles used is 8 with the most optimal paths chosen respectively: (0, 43, 42, 41, 40, 44, 46, 45, 48, 51, 50, 52, 49, 47, 0), (0, 5, 3, 7, 8, 10, 11, 9, 6, 4, 2, 1, 75, 0), (0, 32, 33, 31, 35, 37, 38, 39, 36, 34, 0), (0, 67, 65, 63, 62, 74, 72, 61, 64, 68, 66, 69, 0), (0, 20, 24, 25, 27, 29, 30, 28, 26, 23, 22, 21, 0), (0, 57, 55, 54, 53, 56, 58, 60, 59, 0), (0, 13, 17, 18, 19, 15, 16, 14, 12, 0) and (0, 71, 70, 73, 0)

The final results of carbon emissions, total cost and customer satisfaction are 54.96, 10639.71 and 100 percent respectively. Figures 7 and 8 showcase the comparison between [30] and this paper and the route distribution of the vehicles

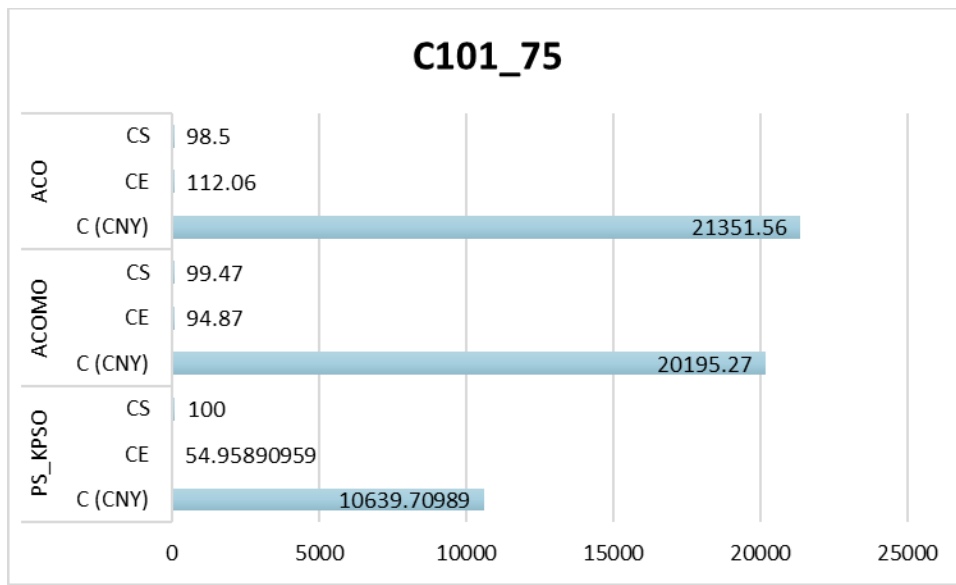


Figure 7: c101(75) comparison with [30] (ACOMO) and ACO+K-Means Clustering (PS_KPSO)

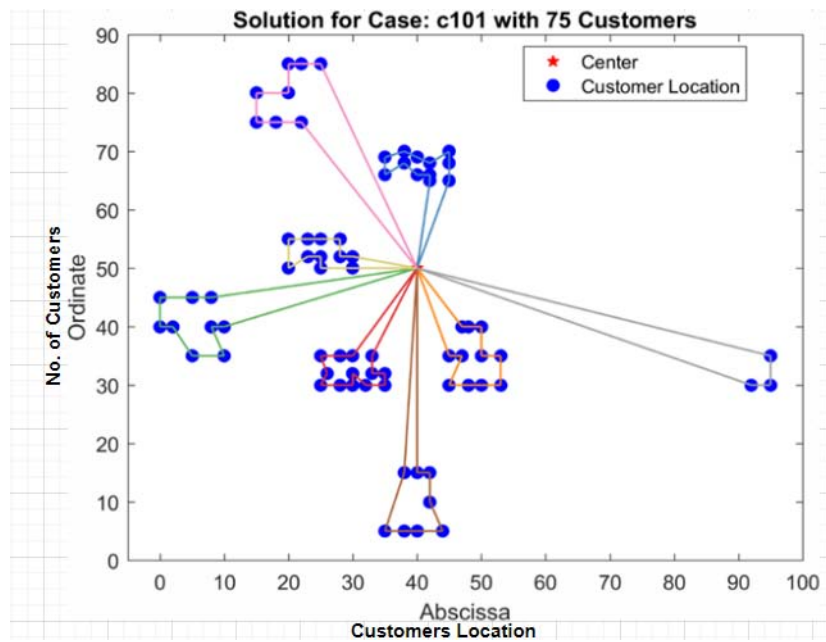


Figure 8: c101(75) vehicle distribution route

iii. 100 Customers

This section has used the c101 (100) dataset. Now looking [30], there are better results in terms of carbon emission, cost and customer satisfaction (69.03, 13561.41 and 100 percent). Instead of 23 vehicles, 10 vehicles have been employed and the most optimal paths are chosen: (0, 5, 3, 7, 8, 10, 11, 9, 6, 4, 2, 1, 75, 0), (0, 43, 42, 41, 40, 44, 46, 45, 48, 51, 50, 52, 49, 47, 0), (0, 20, 24, 25, 27, 29, 30, 28, 26, 23, 22, 21, 0), (0, 67, 65, 63, 62, 74, 72, 61, 64, 68, 66, 69, 0), (0, 90, 87, 86, 83, 82, 84, 85, 88, 89, 91, 0), (0, 57, 55, 54, 53, 56, 58, 60, 59, 0), (0, 98, 96, 95, 94, 92, 93, 97, 100, 99, 0), (0, 32, 33, 31, 35, 37, 38, 39, 36, 34, 0), (0, 13, 17, 18, 19, 15, 16, 14, 12, 0), (0, 81, 78, 76, 71, 70, 73, 77, 79, 80, 0). Figures 9 and 10 showcase the comparison

between [30] and this paper and the route distribution of the vehicles.

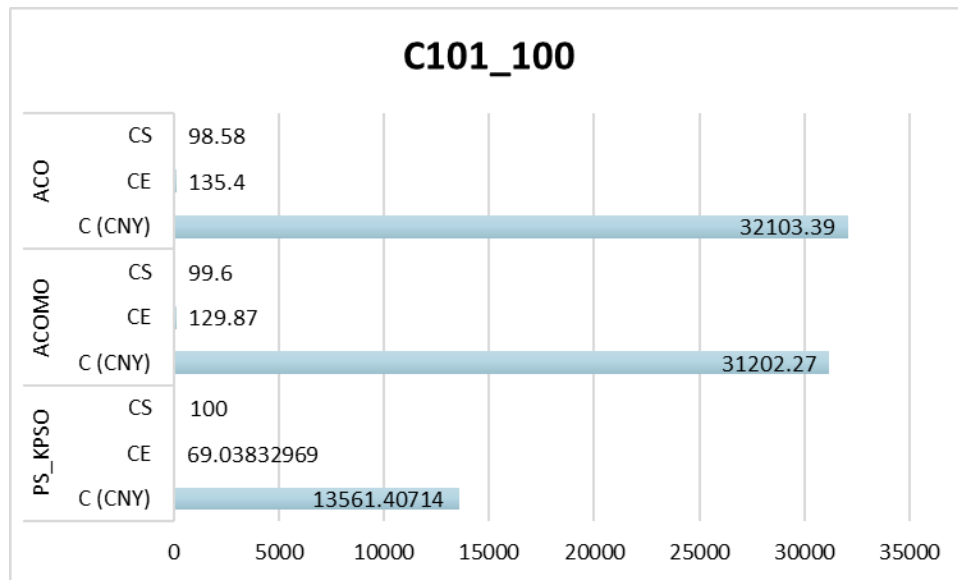


Figure 9: c101(100) comparison with [30] (ACOMO) and ACO+K-Means Clustering (PS_KPSO)

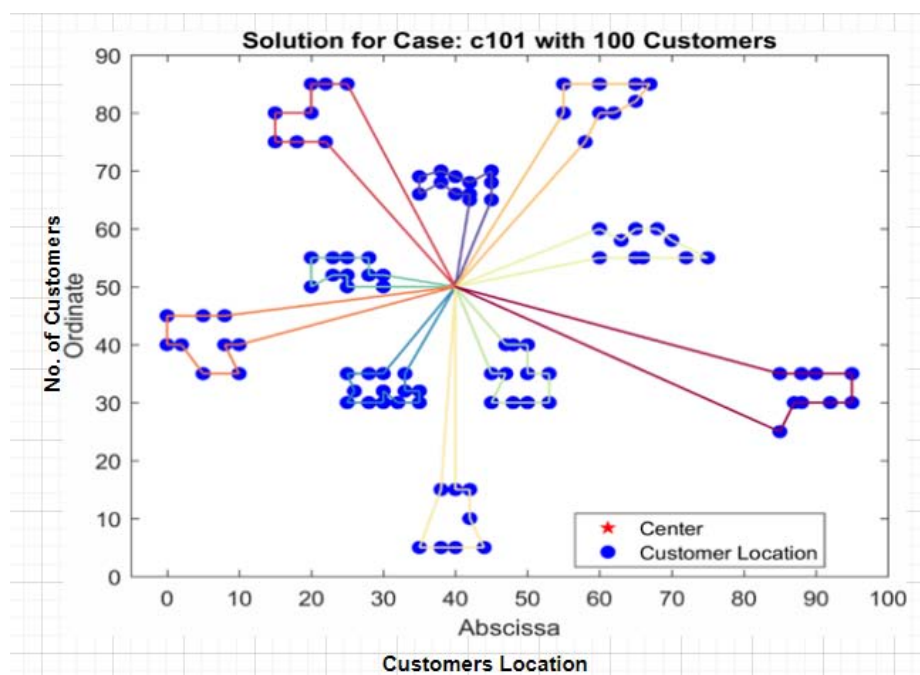


Figure 10: c101(100) vehicle distribution route

d) Comparison

Looking at all the results above, it is easily discernible that the ACO+K-Means clustering algorithm has performed way better than the improved Ant Colony algorithm and the normal Ant Colony Algorithm. With lesser number of vehicles employed, lesser carbon emission levels noted and better cost management, the proposed system has shown its effectiveness and viability for usage in the real-world logistics problems. The proposed algorithm PS_KPSO has provided about 10.37%, 46.9%, 61.98% and 78.81% reduction in total

costs for 25, 50, 75 and 100 customers while there are about 46.61% , 53.27% and 61.16% reduction in total carbon emissions for 50, 75 and 100 customers, when compared with [30]. Along with the aforementioned improvements, there is 100% customer satisfaction in all the cases. The proposed algorithm (ACO+K-Means Clustering) has outperformed the Modified Ant Colony Algorithm and the original Ant Colony algorithm. Table 3 show results compared with Table of that obtained by the proposed algorithm and modified ant colony algorithm.

Table 3: Comparison between algorithms [30]

| PID | NUM_CUST | ACOMO | | | | PS_KPSO | | | |
|----------|----------|----------|--------|-------|----|-------------|-------------|-----|----|
| | | C (CNY) | CE | CS | NV | C (CNY) | CE | CS | NV |
| C101_25 | 25 | 3481.31 | 19.73 | 98.4 | 5 | 3138.070823 | 19.50420315 | 100 | 3 |
| C101_50 | 50 | 9583.7 | 53.75 | 99.2 | 11 | 5942.717648 | 33.43172956 | 100 | 5 |
| C101_75 | 75 | 20195.27 | 94.87 | 99.47 | 17 | 10639.70989 | 54.95890959 | 100 | 8 |
| C101_100 | 100 | 31202.27 | 129.87 | 99.6 | 23 | 13561.40714 | 69.03832969 | 100 | 10 |

e) Results with other test cases

In the Solomon-100 dataset, there are three formats of destination grouping. One is a cluster format (C), one is a random format (R) and one is a random-clustered format (RC). These three formats have been

used for 25, 50, 75 and 100 customers. So other than C101, there are C201, R211, R201 and RC201. The comparison between the proposed algorithm (ACO+K-Means algorithm) and modified Ant Colony algorithm [30] have been given in Table 4.

Table 4: Remaining test set results comparison

| PID | NUM_CUST | TOT_VEH | PS_KACO | | | | ACOMO | | | |
|------------|----------|---------|---------|---------|---------|-----|-------|---------|--------|-----|
| | | | NV | C | CE | CS | NV | C | CE | CS |
| C_201_25 | 25 | 25 | 2 | 215.543 | 14.864 | 100 | 3 | 613.81 | 22.28 | 100 |
| C_201_50 | 50 | 25 | 2 | 444.961 | 19.7345 | 100 | 6 | 1232.8 | 42.46 | 100 |
| C_201_75 | 75 | 25 | 3 | 511.09 | 26.2824 | 100 | 9 | 2177.34 | 58.81 | 100 |
| C_201_100 | 100 | 25 | 3 | 591.557 | 27.9907 | 100 | 13 | 2221.71 | 99.08 | 100 |
| r_201_25 | 25 | 25 | 2 | 543.693 | 21.8306 | 100 | 5 | 946.39 | 45.17 | 100 |
| r_201_50 | 50 | 25 | 2 | 1039.39 | 32.3543 | 100 | 6 | 1404.82 | 69.75 | 100 |
| r_201_75 | 75 | 25 | 3 | 1368.58 | 44.4869 | 100 | 7 | 2482.75 | 98.4 | 100 |
| r_201_100 | 100 | 25 | 4 | 1995.19 | 62.9339 | 100 | 13 | 2931.63 | 111.02 | 100 |
| r_211_25 | 25 | 25 | 1 | 375.432 | 13.1144 | 100 | 2 | 400 | 24.11 | 100 |
| r_211_50 | 50 | 25 | 2 | 1391.42 | 39.8279 | 100 | 7 | 600 | 44.87 | 100 |
| r_211_75 | 75 | 25 | 2 | 1199.99 | 35.7638 | 100 | 7 | 873.41 | 72.94 | 100 |
| r_211_100 | 100 | 25 | 3 | 1867.28 | 55.0745 | 100 | 9 | 1080.64 | 84.49 | 100 |
| r_c201_25 | 25 | 25 | 2 | 454.046 | 19.9274 | 100 | 4 | 847.18 | 31.43 | 100 |
| r_c201_50 | 50 | 25 | 3 | 974.703 | 36.1249 | 100 | 8 | 1554.47 | 80 | 100 |
| r_c201_75 | 75 | 25 | 4 | 1623.5 | 55.0429 | 100 | 10 | 2186.5 | 121.39 | 100 |
| r_c201_100 | 100 | 25 | 4 | 1927.47 | 61.4963 | 100 | 11 | 2959.41 | 139.2 | 100 |

The data from Table 4 helps in evaluating the effectiveness of the proposed algorithm. Even with increase in the number of customers, be it clustered, random or both, there is barely any increase in the number of vehicles employed. With an average of 2.625 vehicles per case, this greatly affects the total travel, storage, damage and fuel costs while reducing the carbon footprint by a great extent, ultimately helping not only the economy of the organisation but also trying to improve the environmental condition of the Earth. It can be assumed from the results data that there is a high probability of increase in number of customers. As the number of vehicles employed is less, there is scope of increasing customer reach and maybe there is a chance of increasing the speed of delivery. With the new electronic vehicle usage, there will be even more cuts in the carbon footprint value and better customer coverage.

f) Terms

The below terms explain the pareto optimal solution which are explained from figure 3 to figure 24. It explains the co-efficient used in capacity of the vehicles and number of customers.

C101(25)- Cost efficient Pareto optimal solution set and its multidimension interpolation for 25 customers

C101(50)- Cost efficient Pareto optimal solution set and its multidimension interpolation for 50 customers) Cost efficient Pareto optimal solution set and its multidimension interpolation for 75 customers

R101(25)- Fuel Co efficient solution set and its multidimension interpolation for 25 customers

R101(50)- Fuel Co efficient solution set and its multidimension interpolation for 55 customers

R101(75)- Fuel Co efficient solution set and its multidimension interpolation for 75 customers

RC101(25)- Fuel and cost Co efficient solution set and its multidimension interpolation for 25 customers
 RC101(50)- Fuel and cost Co efficient solution set and its multidimension interpolation for 50 customers
 RC101(75)- Fuel and cost Co efficient solution set and its multidimension interpolation for 75 customers

The Figures 11 to figure 24 shows case 1 and case 2 shows optimal path, cost and fuel efficiency. shows the obtained Pareto optimal solution and its multidimensional linear interpolation by using the c101 (25) to c(101)100 customers test table for the above table 4.

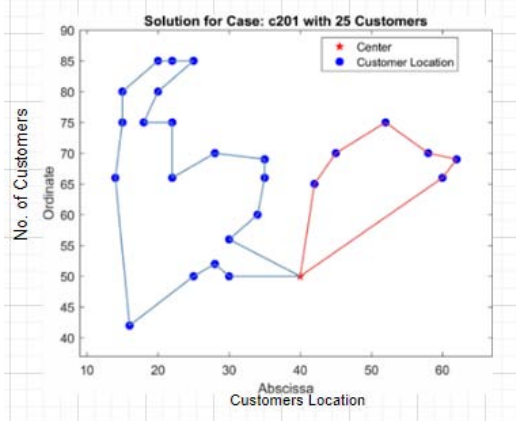


Figure 11: c201(25) vehicle distribution route

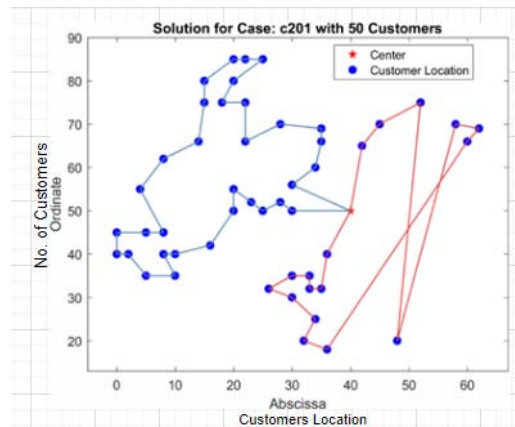


Figure 12.: c201(50) vehicle distribution route

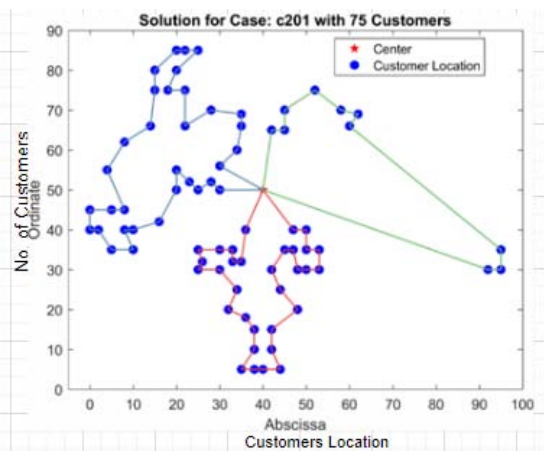


Figure 13.: c201(75) vehicle distribution route

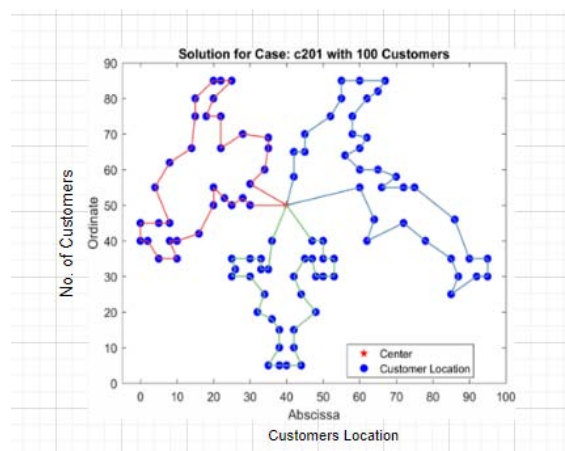


Figure 14.: c201(100) vehicle distribution route

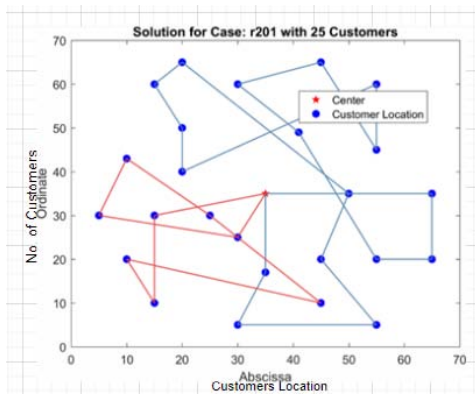


Figure 15.: r201(25) vehicle distribution route

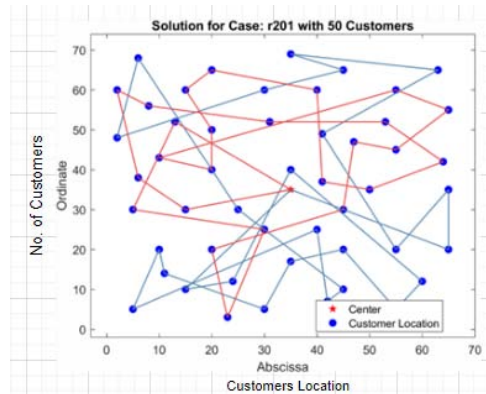


Figure 16.: r201(50) vehicle distribution route

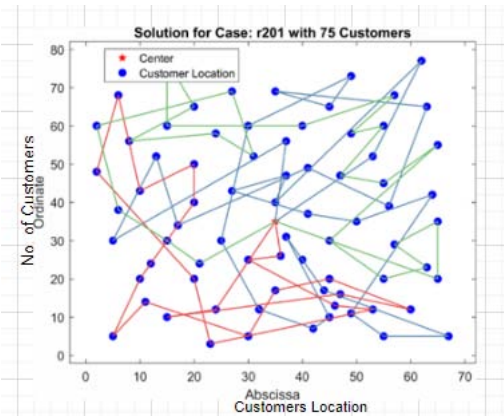


Figure 17.: r201(75) vehicle distribution route

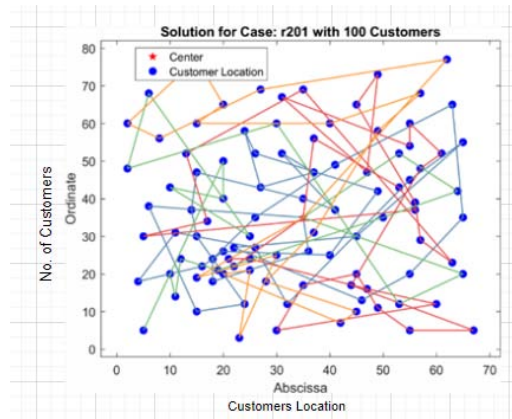


Figure 18.: r201(100) vehicle distribution route

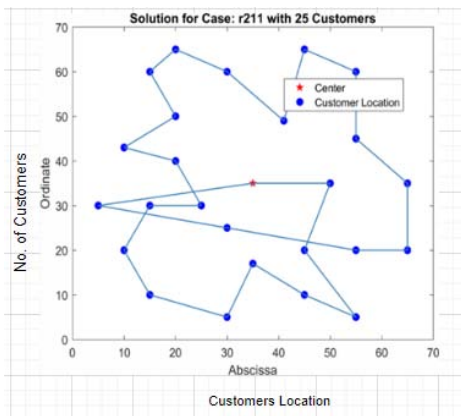


Figure 19.: r211(25) vehicle distribution route

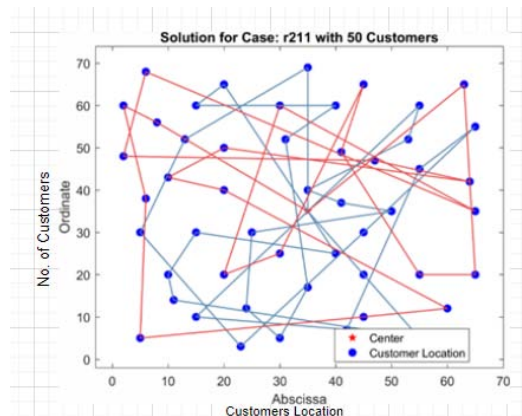


Figure 20.: r211(50) vehicle distribution route

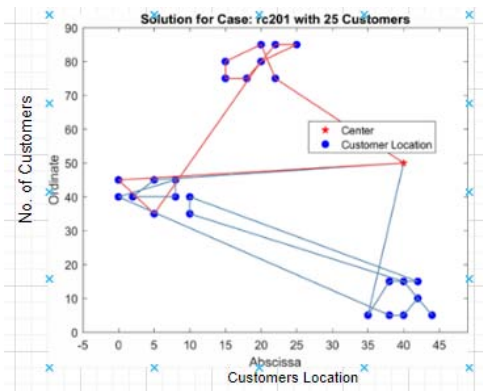


Figure 21.: rc201(25) vehicle distribution route

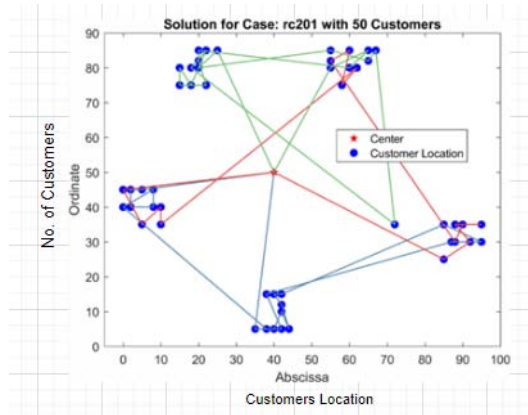


Figure 22.: rc201(50) vehicle distribution route



Figure 23.: rc201(75) vehicle distribution route

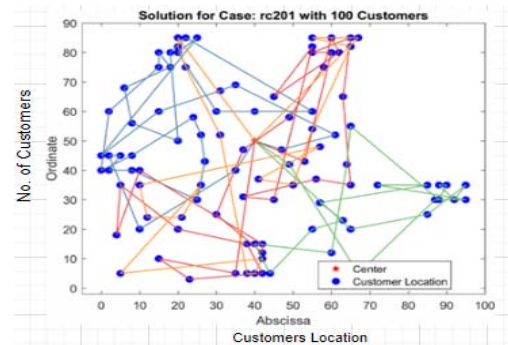


Figure 24.: rc201(100) vehicle distribution route

ptimal solutions in the figure. The number of optimal solutions is less than the one with the test set having 25 to 100 customers. Generally, the fewer the optimal solutions there are, the fewer distribution plans the distribution center can choose. As an important factor in cold chain logistics, temperature control plays an active role in cold chain logistics distribution. The strict

temperature control can effectively reduce the impact of temperature fluctuations on cargo in cold chain transportation. To a certain extent, refrigeration equipment has caused an increase in carbon emission. Below table 5 explains from optimal path, fuel, and cost efficiency for above figure 11 to figure 24

Table 5: Vehicle distribution route for optimal path, fuel, and cost efficiency

| Figure Number | Comments |
|---|---|
| Figure 11.: c201(25) vehicle distribution route | The Figure 11 shows vehicle distribution for case 2, it shows optimal path for 25 customers |
| Figure 12.: c201(50) vehicle distribution route | The Figure 12 shows vehicle distribution for case 2, it shows optimal path for 50 customers |
| Figure 13.: c201(75) vehicle distribution route | The Figure 13 shows vehicle distribution for case 2, it shows optimal path for 75 customers |
| Figure 14.: c201(100) vehicle distribution route | The Figure 14 shows vehicle distribution for case 2, it shows optimal path for 100 customers |
| Figure 15.: r201(25) vehicle distribution route | The Figure 15 shows vehicle distribution for case 2, it shows fuel efficiency for 25 customers |
| Figure 16.: r201(50) vehicle distribution route | The Figure 16 shows vehicle distribution for case 2, it shows fuel efficiency for 50 customers. |
| Figure 17.: r201(75) vehicle distribution route | The Figure 17 shows vehicle distribution for case 2, it shows fuel efficiency for 75 customers. |
| Figure 18.: r201(100) vehicle distribution route | The Figure 18 shows vehicle distribution for case 2, it shows fuel efficiency for 100 customers. |
| Figure 19.: r211(25) vehicle distribution route | The Figure 19 shows vehicle distribution for case 2, it shows fuel efficiency for 25 customers for random vehicles. |
| Figure 20.: r211(50) vehicle distribution route | The Figure 20 shows vehicle distribution for case 2, it shows fuel efficiency for 50 customers for random vehicles. |
| Figure 21.: rc201(25) vehicle distribution route | The Figure 21 shows vehicle distribution for case 2, it shows fuel and cost efficiency for 25 customers for random vehicles. |
| Figure 22.: rc201(50) vehicle distribution route | The Figure 22 shows vehicle distribution for case 2, it shows fuel and cost efficiency for 50 customers for random vehicles. |
| Figure 23.: rc201(75) vehicle distribution route | Figure 23 shows vehicle distribution for case 2, it shows fuel and cost efficiency for 75 customers for random vehicles. |
| Figure 24.: rc201(100) vehicle distribution route | The Figure 24 shows vehicle distribution for case 2, it shows fuel and cost efficiency for 100 customers for random vehicles. |

VI. CONCLUSION

This paper discusses the vehicle routing problem with time window constraint (VRPTW) along with added constraints of number of vehicles, logistics cost, overall carbon emission rate along with multiple pickup and delivery points. A meta heuristic Ant Colony Algorithm with K-Means Clustering was employed to solve the problem statement. Looking at the literature survey in this paper, it is observable that Vehicle Routing Problem has had several approaches with varying results, which in turn leads to the fact that VRP with added constraints is a difficult problem to solve.

The solution provided in this paper has been compared with [30], which has a similar problem statement, and the results of the proposed ant colony algorithm with K-Meansclustering has performed far better and has provided very less scope of improvement in the discussed problem areas.

In future research on similar topics, it's a hope that this paper will be a good leverage for the researchers and this solution can be further modified for more improvements.

REFERENCES RÉFÉRENCES REFERENCIAS

- Dantzig, G.B., and Ramser, J.H., 1959, "The truck dispatching problem," *Management Science*, Vol. 6, No. 1, pp. 80-91.
- Clarke, G., and Wright, J.W., 1964, "Scheduling of vehicles from a central depot to a number of delivery points," *Operations Research*, Vol. 12, No. 4, pp. 568-581.
- MALANDRAKI C, DASKIN M S. Time dependent vehicle routing problems: formulations, properties and heuristic algorithms. *Transportation Science*, 1992, 26(3): 185 – 200.
- ICHOUA S, GENDREAU M, POTVIN J Y. Vehicle dispatching with time-dependent travel times. *European Journal of Operational Research*, 2003, 144(2): 379 – 396.
- OSVALD A, STIRN L Z. A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food. *Journal of Food Engineering*, 2008, 85(2): 285 – 295.
- KUO Y. Using simulated annealing to minimize fuel consumption for the time-dependent vehicle routing problem. *Computers & Industrial Engineering*, 2010, 59(1): 157 – 165.
- BALSEIRO S R, LOISEAU I, RAMONET J. An ant colony algorithm hybridized with insertion heuristics for the time dependent vehicle routing problem with time windows. *Computers & Operations Research*, 2011, 38(6): 954 – 966.
- HASHIMOTO H, YAGIURA M, IBARAKI T. An iterated local search algorithm for the time-dependent vehicle routing problem with time windows. *Discrete Optimization*, 2008, 5(2): 434 – 456.
- FIGLIOZZI M A. The time dependent vehicle routing problem with time windows: benchmark problems, an efficient solution algorithm, and solution characteristics. *Transportation*
- KOK A L, HANS E, SCHUTTEN J. Vehicle routing under time-dependent travel times: the impact of congestion avoidance. *Computers & Operations Research*, 2012, 39(5): 910 – 918.
- MINOCHA B, TRIPATHI S. Two phase algorithm for solving VRPTW problem. *International Journal of Artificial Intelligence and Expert Systems*, 2013, 4(1): 1 – 16.
- Müller, J., 2010, "Approximative Solutions to the Bicriterion Vehicle Routing Problem with Time Windows", *European Journal of Operational Research*, 202, pp. 223–231
- Kumar, S. N., and Panneerselvam, R., 2012, "A Survey on the Vehicle Routing Problem and Its Variants", *Intelligent Information Management*, 4, pp. 66-74
- H. Mao, J. Shi, Y. Zhou and G. Zhang, "The Electric Vehicle Routing Problem With Time Windows and Multiple Recharging Options," in *IEEE Access*, vol. 8, pp. 114864-114875, 2020, doi: 10.1109/ACCESS.2020.3003000.
- Y. Wu, B. Zheng and X. Zhou, "A Disruption Recovery Model for Time-Dependent Vehicle Routing Problem With Time Windows in Delivering Perishable Goods," in *IEEE Access*, vol. 8, pp. 189614-189631, 2020, doi: 10.1109/ACCESS.2020.3032018.
- L. Wu, Z. He, Y. Chen, D. Wu and J. Cui, "Brainstorming-Based Ant Colony Optimization for Vehicle Routing With Soft Time Windows," in *IEEE Access*, vol. 7, pp. 19643-19652, 2019, doi: 10.1109/ACCESS.2019.2894681.
- Bowerman, R.L.; Calamai, P.H. The space filling curve with optimal partitioning heuristic for the vehicle routing problem. *Eur. J. Oper. Res.* 1994, 76, 128–142.
- Dondo, R.; Cerda, . A cluster-based optimization approach for the multi depot heterogeneous fleet vehicle routing problem with time windows. *Eur. J. Oper. Res.* 2007, 176, 478–1507.
- Cordeau, .F.; Gendreau, M.; Laporte, G.; Potvinand, .Y.; Semet, F. A guide to vehicle routing heuristics. *J. Oper. Res. Soc.* 2002, 53, 512–522.
- Szeto, W.Y.; Wu, Y.; Ho, S.C. An artificial ee colony algorithm for the capacitated vehicle routing problem. *Eur. .Oper. Res.* 2011, 215, 126–135.
- Bullnheimer, B.; Hartl, R.F.; Strauss, C. Applying the ant system to the vehicle routing problem. In *Meta-Heuristics: advances and Trends in Local Search or Optimization*; Voss, S., Martello, S., Osman, I.H.,

- Roucairol, C., Eds.; Kluwer Academic Publishers: Boston, MA, USA, 2012; pp. 285–296.
22. Tan, W.F.; Lee, L.S.; Majid, Z.A.; Seow, H.V. Ant colony optimization for capacitated vehicle routing problem. *J. Comput. Sci.* 2012, **8**, 846–852.
 23. Doerner, K.; Gronalt, M.; Hartl, R.F.; Reimann, M.; Strauss, C.; Stummer, M. Savings Ants for the vehicle routing problem. *Lect. Notes Comput. Sci.* 2002, **2279**, 73–109.
 24. Ma, Y.; Han, .; Kang, K.; Yan, F. An Improved ACO for the multi-depot vehicle routing problem with time windows. In *Proceedings of the Tenth International Conference on Management Science and Engineering Management*; Xu, ., Hajiyev, A., Nickel, S., Gen, M., Eds.; Springer: Singapore, 2018; pp. 1181–1189.
 25. Zhang, X.; Tang, L. A new hybrid ant colony optimization algorithm for the vehicle routing problem. *Pattern Recognit. Lett.* 2009, **30**, 848–855. [CrossRef]
 26. Mazzeo, S.; Loiseau, I. An ant colony algorithm for the capacitated vehicle routing. *Electron. Notes Discret. ath.* 2004, **18**, 181–186. [CrossRef]
 27. Yu, B.; Yang, Z.Z.; Yao, B.Z. An improved ant colony optimization for vehicle routing problem. *Eur. Oper. Res.* 2009, **196**, 171–176. [CrossRef]
 28. Ma, Y.; Han, .; Kang, K.; Yan, F. An Improved ACO for the multi-depot vehicle routing problem with time windows. In *Proceedings of the Tenth International Conference on Management Science and Engineering Management*; Xu, ., Hajiyev, A., Nickel, S., Gen, M., Eds.; Springer: Singapore, 2018; pp. 1181–1189.
 29. Bullnheimer, B.; Hartl, R.F.; Strauss, C. An improved ant system for the vehicle routing problem. *Ann. Oper. Res.* 1999, **89**, 319–328.
 30. B. Zhao, H. Gui, H. Li and J. Xue, "Cold Chain Logistics Path Optimization via Improved Multi-Objective Ant Colony Algorithm," in *IEEE Access*, vol. 8, pp. 142977-142995, 2020, doi: 10.1109/ACCESS.2020.3013951.



GLOBAL JOURNALS GUIDELINES HANDBOOK 2022

WWW.GLOBALJOURNALS.ORG

MEMBERSHIPS

FELLOWS/ASSOCIATES OF COMPUTER SCIENCE RESEARCH COUNCIL FCSRC/ACSRC MEMBERSHIPS

INTRODUCTION



FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.



BENEFIT

TO THE INSTITUTION

GET LETTER OF APPRECIATION

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.



EXCLUSIVE NETWORK

GET ACCESS TO A CLOSED NETWORK

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly. They should also be open to reaching out by other.

Career

Credibility

Exclusive

Reputation



CERTIFICATE

CERTIFICATE, LOR AND LASER-MOMENTO

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career

Credibility

Exclusive

Reputation



DESIGNATION

GET HONORED TITLE OF MEMBERSHIP

Fellows can use the honored title of membership. The "FCSRC" is an honored title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career

Credibility

Exclusive

Reputation

RECOGNITION ON THE PLATFORM

BETTER VISIBILITY AND CITATION

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career

Credibility

Reputation

FUTURE WORK

GET DISCOUNTS ON THE FUTURE PUBLICATIONS

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career

Financial



GJ ACCOUNT

UNLIMITED FORWARD OF EMAILS

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career

Credibility

Reputation



PREMIUM TOOLS

ACCESS TO ALL THE PREMIUM TOOLS

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

CONFERENCES & EVENTS

ORGANIZE SEMINAR/CONFERENCE

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career

Credibility

Financial

EARLY INVITATIONS

EARLY INVITATIONS TO ALL THE SYMPOSIUMS, SEMINARS, CONFERENCES

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive





PUBLISHING ARTICLES & BOOKS

EARN 60% OF SALES PROCEEDS

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive

Financial

REVIEWERS

GET A REMUNERATION OF 15% OF AUTHOR FEES

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

ACCESS TO EDITORIAL BOARD

BECOME A MEMBER OF THE EDITORIAL BOARD

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career

Credibility

Exclusive

Reputation

AND MUCH MORE

GET ACCESS TO SCIENTIFIC MUSEUMS AND OBSERVATORIES ACROSS THE GLOBE

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.



BENEFIT

TO THE INSTITUTION

GET LETTER OF APPRECIATION

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.



EXCLUSIVE NETWORK

GET ACCESS TO A CLOSED NETWORK

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly. They should also be open to reaching out by other.

Career

Credibility

Exclusive

Reputation



CERTIFICATE

CERTIFICATE, LOR AND LASER-MOMENTO

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career

Credibility

Exclusive

Reputation



DESIGNATION

GET HONORED TITLE OF MEMBERSHIP

Associates can use the honored title of membership. The "ACSRC" is an honored title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career

Credibility

Exclusive

Reputation

RECOGNITION ON THE PLATFORM

BETTER VISIBILITY AND CITATION

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career

Credibility

Reputation

FUTURE WORK

GET DISCOUNTS ON THE FUTURE PUBLICATIONS

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career

Financial



GJ ACCOUNT

UNLIMITED FORWARD OF EMAILS

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career

Credibility

Reputation



PREMIUM TOOLS

ACCESS TO ALL THE PREMIUM TOOLS

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

CONFERENCES & EVENTS

ORGANIZE SEMINAR/CONFERENCE

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career

Credibility

Financial

EARLY INVITATIONS

EARLY INVITATIONS TO ALL THE SYMPOSIUMS, SEMINARS, CONFERENCES

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive





PUBLISHING ARTICLES & BOOKS

EARN 30-40% OF SALES PROCEEDS

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive

Financial

REVIEWERS

GET A REMUNERATION OF 15% OF AUTHOR FEES

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

AND MUCH MORE

GET ACCESS TO SCIENTIFIC MUSEUMS AND OBSERVATORIES ACROSS THE GLOBE

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.



| ASSOCIATE | FELLOW | RESEARCH GROUP | BASIC |
|---|---|---|---|
| <p>\$4800 lifetime designation</p> <hr/> <p>Certificate, LoR and Momento 2 discounted publishing/year Gradation of Research 10 research contacts/day 1 GB Cloud Storage GJ Community Access</p> | <p>\$6800 lifetime designation</p> <hr/> <p>Certificate, LoR and Momento Unlimited discounted publishing/year Gradation of Research Unlimited research contacts/day 5 GB Cloud Storage Online Presense Assistance GJ Community Access</p> | <p>\$12500.00 organizational</p> <hr/> <p>Certificates, LoRs and Momentos Unlimited free publishing/year Gradation of Research Unlimited research contacts/day Unlimited Cloud Storage Online Presense Assistance GJ Community Access</p> | <p>APC per article</p> <hr/> <p>GJ Community Access</p> |



PREFERRED AUTHOR GUIDELINES

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from <https://globaljournals.org/Template.zip>

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

BEFORE AND DURING SUBMISSION

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct*, along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s) names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

POLICY ON PLAGIARISM

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures



- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

AUTHORSHIP POLICIES

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

PREPARING YOUR MANUSCRIPT

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.



Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

- a) A title which should be relevant to the theme of the paper.
- b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
- c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
- d) An introduction, giving fundamental background objectives.
- e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
- f) Results which should be presented concisely by well-designed tables and figures.
- g) Suitable statistical data should also be given.
- h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

- i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
- j) There should be brief acknowledgments.
- k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.



FORMAT STRUCTURE

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in a research paper?" Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.



Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

PREPARATION OF ELETRONIC FIGURES FOR PUBLICATION

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

TIPS FOR WRITING A GOOD QUALITY COMPUTER SCIENCE RESEARCH PAPER

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.



6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10. Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.



20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium through which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

INFORMAL GUIDELINES OF RESEARCH PAPER WRITING

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.



Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.



The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.



Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."



Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

THE ADMINISTRATION RULES

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.



CRITERION FOR GRADING A RESEARCH PAPER (COMPILATION)
BY GLOBAL JOURNALS INC. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

| Topics | Grades | | |
|-------------------------------|--|---|--|
| | A-B | C-D | E-F |
| <i>Abstract</i> | Clear and concise with appropriate content, Correct format. 200 words or below | Unclear summary and no specific data, Incorrect form Above 200 words | No specific data with ambiguous information Above 250 words |
| <i>Introduction</i> | Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited | Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter | Out of place depth and content, hazy format |
| <i>Methods and Procedures</i> | Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads | Difficult to comprehend with embarrassed text, too much explanation but completed | Incorrect and unorganized structure with hazy meaning |
| <i>Result</i> | Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake | Complete and embarrassed text, difficult to comprehend | Irregular format with wrong facts and figures |
| <i>Discussion</i> | Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited | Wordy, unclear conclusion, spurious | Conclusion is not cited, unorganized, difficult to comprehend |
| <i>References</i> | Complete and correct format, well organized | Beside the point, Incomplete | Wrong format and structuring |



INDEX

A

Aesthetic · 17
Anatomical · 10
Annealing · 26, 27, 37

B

Breached · 14

C

Combat · 9
Convolutions · 4, 19
Corpus · 4
Cumbersome · 2, 3

D

Distortion · 16, 17, 19, 20, 21, 22, 23, 40

E

Elitist · 29, 30
Emissions · 24, 26, 30, 31, 32, 33
Eradicated · 9

G

Gradient · 17, 18, 19

H

Heuristics · 25, 26, 27, 37, 38

I

Infractions · 14
Infrared · 10
Ingenious · 12

P

Parsing · 2, 3, 4
Peculiarity · 28
Pertaining · 1
Precision · 6, 13, 14
Prevalent · 1
Pristine · 16, 18, 19

R

Recursion · 26

S

Segregation · 9
Semantic · 2, 3, 4
Siamese · 18
Stemming · 2
Stochastic · 19
Syntactic · 2



save our planet



Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org



ISSN 9754350