IAES International Journal of Artificial Intelligence (IJ-AI)

Vol 12, No 1: March 2023

IAES International Journal of Artificial Intelligence (IJ-AI)

IAES International Journal of Artificial Intelligence (IJ-AI), ISSN/e-ISSN 2089-4872/2252-8938 publishes articles in the field of artificial intelligence (AI). The scope covers all artificial intelligence (AI) and machine learning (ML) areas and theirs applications in the following topics: neural networks; fuzzy logic; simulated biological evolution algorithms (like genetic algorithm, ant colony optimization, etc); reasoning and evolution; intelligence applications; computer vision and speech understanding; multimedia and cognitive informatics, data mining and machine learning tools, heuristic and AI planning strategies and tools, computational theories of learning; technology and computing (like particle swarm optimization); intelligent system architectures; knowledge representation; bioinformatics; natural language processing; multiagent systems; supervised learning; unsupervised learning; deep learning; big data and AI approaches; reinforcement learning; and learning with generative adversarial networks; etc. This journal is indexed in Scopus and all published papers since 2018 issues were included in scopus.com.

Focus and Scope

The IAES International Journal of Artificial Intelligence (IJ-AI), ISSN/e-ISSN 2089-4872/2252-8938 covers all topics of artificial intelligence and soft computing and their applications, including but not limited to:

- Neural networks
- Reasoning and evolution
- Intelligent search
- Intelligent planning
- Intelligence applications
- Computer vision and speech understanding
- Multimedia and cognitive informatics
- Data mining and machine learning tools, heuristic and AI planning strategies and tools, computational theories of learning
- Technology and computing (like particle swarm optimization); intelligent system architectures
- Knowledge representation
- Bioinformatics
- Natural language processing
- Automated reasoning
- Logic programming
- Machine learning
- Visual/linguistic perception
- Evolutionary and swarm algorithms
- Derivative-free optimisation algorithms
- Fuzzy sets and logic
- Rough sets
- Simulated biological evolution algorithms (like genetic algorithm, ant colony optimization, etc)
- Multi-agent systems
- Data and web mining
- Emotional intelligence
- Hybridisation of intelligent models/algorithms
- Parallel and distributed realisation ofintelligent algorithms/systems
- Application in pattern recognition, image understanding, control, robotics and bioinformatics
- Application in system design, system identification, prediction, scheduling and game playing
- Application in VLSI algorithms and mobile communication/computing systems

Principal Contact
Prof. Dr. Eugene Yu-Dong Zhang
Editor-in-Chief, IJ-AI
Chair in Knowledge Discovery and Machine Learning
Associate Fellow of Higher Education Academy
IEEE Senior Member
ACM Senior Member

Contact
F26 Informatics Building
Department of Informatics
University of Leicester, University Road,
Leicester, LE1 7RH, UK
Email: ijai@iaesjournal.com

IAES International Journal of Artificial Intelligence (IJ-AI)

Editorial Team

IAES International Journal of Artificial Intelligence (IJ-AI)

Vol. 12, No. 1 March 2023

Table of Contents

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 1~11
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp1-11

Automatic identification system-based trajectory clustering framework to identify vessel movement pattern

I Made Oka Widyantara[1], I Putu Noven Hartawan[2], Anak Agung Istri Ngurah Eka Karyawati[3], Ngurah Indra Er[1], Ketut Buda Artana[4]
[1]Department of Electrical Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia
[2]Postgraduate Program of Electrical Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia
[3]Department of Computer Science, Faculty of Math and Natural Science, Udayana University, Bali, Indonesia
[4]Department of Marine Engineering, Faculty of Marine Technology, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

Article Info

Article history:

Received Nov 20, 2021
Revised Jul 20, 2022
Accepted Aug 18, 2022

Keywords:

Automatic identification system
Data mining
Density-based spatial clustering of applications with noise
Longest common subsequence
Trajectory
Vessel

ABSTRACT

Automatic identification system (AIS) is a vessel radio navigation equipment that has been determined by international maritime organization (IMO). Historical AIS data can be utilized for anomaly detection, trajectory prediction, and vessel trajectory planning. These benefits can be achieved by identifying the vessel's trajectory pattern through trajectory clustering. However, more effort is needed in trajectory clustering using AIS data due to their large volume and the significant number of deficiencies. In addition, trajectory clustering cannot be directly applied to trajectory data, which also applies to vessel trajectory. Therefore, we propose a trajectory clustering framework by combining douglas peucker (DP), longest common subsequence (LCSS), multi-dimensional scaling (MDS), and density-based spatial clustering of applications with noise (DBSCAN). Our experiments, carried out with AIS data for the Lombok Strait, Indonesia, showed that the trajectory compression with DP significantly accelerates the similarity measurement process. Moreover, we found that the LCSS is the optimal algorithm for similarity measurement of vessel trajectories based on AIS data. We also applied the right combination of MDS and DBSCAN in density-based clustering. The proposed framework can distinguish trajectoriess in different directions, identify the noise, and produce good quality clusters in relatively fast total processing time.

Corresponding Author:

I Made Oka Widyantara
Department of Electrical Engineering, Faculty of Engineering, Udayana University
Kampus Unud Road, Jimbaran, Bali, Indonesia
Email: oka.widyantara@unud.ac.id

1. INTRODUCTION

The automatic identification system (AIS) is a radio navigation device that uses very high frequency (VHF) to transmit vessel data automatically between vessels at sea and receivers on land. Every vessel over 300 gross tons (GT) must have an AIS signal transmitter, according to the international maritime organization (IMO) regulation [1]–[4]. Vessel location, speed, lane, direction, turn rate, destination, and expected time of arrival are among the dynamic data supplied by AIS. Static data as are vessel name, vessel maritime mobile service identity (MMSI ID), message identity (ID), vessel type, vessel size, and current time also provided. Furthermore, AIS data has the advantage of providing the highest volume of vessel position data with wide water area coverage [5] and commercially accessible or open-source ais data, which other vessel reporting systems do not have [6]. Many things may be evaluated using AIS data due to the vast

number of data, including traffic analysis, transportation logistics, monitoring, collisions, pollutants, oil spills, and fishing activity [7].

Based on the history of existing AIS data, data mining techniques and artificial intelligence systems can be utilized to identify vessel route pattern at sea. Anomaly detection, trajectory prediction, and vessel trajectory planning can all be done after vessel trajectory pattern has gotten [8]. Moreover, clustering can group vessel trajectories based on the characteristics of each trajectory. The vessel trajectory pattern, whether it is already or not yet known, will appear based on the clusters resulting from the AIS data clustering process [9]–[11]. However, more effort is needed in trajectory clustering using AIS data due to its large volume and usually many deficiencies, such as low data quality, irregular AIS message time intervals, and poor data integrity [12]. The anomalies occur because AIS messages are sent from vessels with various types, delivery distances, and geographical conditions. In addition, trajectory clustering cannot be directly applied to the trajectory data, including the vessel trajectory data from AIS. Inherently, vessel trajectories are different from traditional data commonly used in clustering methods [10]. Therefore, a suitable trajectory clustering framework is needed to generate vessel trajectory patterns using AIS data. The main steps that need to be carried out in AIS data trajectory clustering are data pre-processing, similarity measurement, and the clustering process itself.

Data pre-processing is a crucial phase in data mining, and it also applies to vessel trajectory clustering [13]. The most time-consuming phase in data mining is data-preparation, which will take longer than the main data mining process itself. Incomplete data, noise, data without attributes, and repeating data are all possible to found in real-world scenarios. The length and shape of the vessel's trajectory varies greatly in the AIS data. Moreover, AIS data often has an abnormal trajectory pattern, which will mislead the algorithm used [14].

Trajectory similarity measurement is a determining factor in trajectory clustering [15]. The method used must be able to make the distance between different trajectories as far as possible and the same trajectory as close as possible. Based on previous research, several similarity measurement methods are commonly used in trajectory clustering with AIS data. Research in [16]–[18] applied hausdorff distance (HD) for trajectory clustering, where HD can identify the shape of the trajectory, calculates the maximum shortest distance value from one trajectory to another, and calculates the average value of the two maximums as distance. However, HD is inadequate in identifying the direction of the trajectory due to its sensitivity to noise [15]. HD also has shortcomings in measuring distance in dense water areas, thus giving incorrect cluster results [16]. Li *et al.* [9] applied dynamic time warping (DTW) in trajectory clustering on the bridge area waterway and Mississippi river. Li *et al.* [10] used merge distance (MD) in trajectory clustering on the bridge waterways. Furthermore, Li *et al.* [9] and Li *et al.* [10] conducted clustering with less varied trajectory data. Li *et al.* [10] showed that DTW and MD have the same accuracy, but DTW is superior in terms of processing time, because MD is a more complex algorithm. The shortcoming of DTW is that the resulting distance greatly affects the noise and sampling rate of the track. It is a potential challenge because the AIS data contains redundant vessel positions caused by the vessel sending AIS messages within the span of 3-10 seconds [19].

Based on the research in [20], partition-based methods, hierarchy-based methods, density-based methods, grid-based methods, and model-based methods are the five categories of clustering methods. The following are some previous studies in the context of trajectory clustering. Partition based clustering is a type of clustering method in which the number or center of clusters is identified before processing is applied. K-means and K-medoids are representations of partition-based methods. Li *et al.* [9] utilized K-means as a vessel trajectory clustering method using AIS data. Furthermore, principal component analysis (PCA) is used to find the value of k in the same research. Zhen *et al.* [17] used K-medoids as a clustering method which will later be classified to detect anomalies. However, both methods cannot automatically detect noise. Density based is a clustering method based on point density. Density-based spatial clustering of applications with noise (DBSCAN) is the most frequently used density-based method. Research in [10], [16], [21], used DBSCAN in the vessel trajectory clustering process, where it can automatically search for the number of clusters based on density. DBSCAN can group clusters with irregular shapes and discover noise automatically and effectively [10].

In this study, DBSCAN was chosen as the algorithm for trajectory clustering. DBSCAN is an unsupervised clustering algorithm that does not need the specification of the number of clusters at the beginning [22]. DBSCAN with epsilon (*Eps)* concept is highly dependent on spatial density. However, DBSCAN has a "curse of dimensionality" problem [23]–[25] and to overcome this, our study applies dimensional reduction with the multi-dimensional scaling (MDS) algorithm. MDS is used to reduce the dimensions of the similarity matrix into relative position data which is a low-dimensional representation of the similarity matrix. The MDS data will be utilized and injected into the DBSCAN while the distance from the MDS data will be used to find the optimal epsilon parameter. Furthermore, the similarity measurement

stage uses the longest common subsequence (LCSS) algorithm. LCSS was chosen because it has less effect on noise and different trajectory lengths while can also detect the direction of the trajectory [26]. Douglas peucker (DP) algorithm is proposed at the pre-processing stage to speed up the similarity measurement process. By combining those algorithms into a framework proposed in this study, we can perform trajectory clustering with a good quality and speed from a collection of vessel trajectories based on complex and diverse AIS data, so that the cluster results can be used as a basis for anomaly detection, trajectory prediction, and vessel trajectory planning.

This study uses AIS data in the waters of the Lombok Strait which has the third highest shipping traffic density in Indonesia [27]. The firts stage of the proposed framework are the cleaning of data and the translation of the AIS coordinate data rows into trajectory data. The next stage is to remove unnecessary coordinate points from the vessel's trajectory using DP while also measuring the similarity between existing vessel trajectories using LCSS. The next stage is to change the trajectory similarity distance data from the similarity matrix into spatial points using MDS, and finally to conduct the clustering using DBSCAN. To evaluate the quality of the resulting cluster, a comparison is made between the proposed algorithm and some benchmark algorithms, based on the total time and cluster quality measurements using the silhouette coefficient (SC) method.

2. METHOD

Figure 1 shows an overview of the proposed framework. It starts with raw AIS data containing row coordinates and vessel information based on time. It follows by several processes. The first stage is preprocessing, which includes data cleaning and converting it into vessel trajectory data and then proceed with trajectory compression. Furthermore, each trajectory which is a combination of several coordinates is simplified using DP. After simplifying the trajectory, each trajectory will be measured using LCSS to find the similarity distance between the trajectories. Then, the results of the distance matrix from DTW need to go through a dimension reduction process using MDS to convert three-dimensional (3D) data into two-dimensional (2D) spatial before proceeding to the clustering stage using DBSCAN.

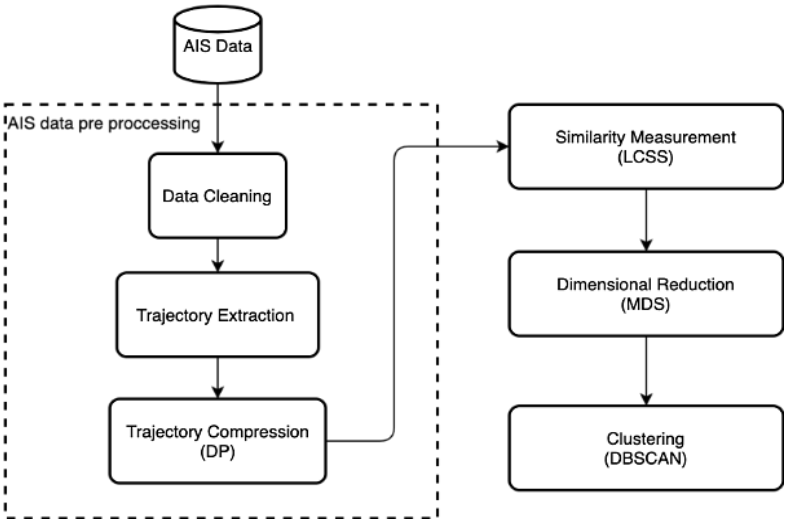


Figure 1. Method overview

2.1. Data preprocessing

Preprocessing is the first step to overcome the problem of AIS data deficiency and make the data ready to be used in trajectory clustering. Base on [13], there are three stages in trajectory clustering preprocessing, there ara cleaning, extraction and compression. At the data cleaning stage, course over ground (COG) and speed over ground (SOG) selections are made. Abnormal data that indicate the vessel is not moving are also eliminated. The SOG selection will also affect the results of the trajectory extraction at a next stage, because it creates a bigger time gap.

Because a vessel can have many trajectories, trajectory extraction cannot be done simply by grouping the vessel's position with MMSI. Therefore, at trajectory extraction phase we trim the trajectory of each MMSI. Referring to research in [18], trajectory trimming is done by measuring the period between

trajectories using time threshold. MMSI markers such as 26XXX591 have been replaced by new markers such as 26XXX591-1, 26XXX591-2, and so on, indicating that the data is a single trajectory unit.

The number of points owned by each vessel's trajectory will make the similarity measure process take a long time. To overcome this, trajectory compression can be done by eliminate the redundant coordinate points without losing the trjectorie's original shape. The algorithm used in the trajectory compression process is the DP algorithm. Due its accuracy and speed, the DP algorithm is widely used in compression of trajectories or moving point objects [28].

2.2. Similarity measurement with LCSS

The measurement of the similarity distance between all trajectories is carried out after performing trajectory compression. LCSS is a well-known method for measuring text similarity, while in the context of trajectory similarity measurement, LCSS can solve the noise problem in trajectory [26]. The main idea of LCSS is to calculate Euclidean distances from several points within two trajectoriess in turn. To solve that, LCSS requires threshold parameters ϵ. When measuring the distance of trajectory, A and B. LCSS consider $a_i(a_i \in A)$ and $b_j(b_j \in B)$ is similar if the distance between the trajectories is less than ϵ and LCSS will ignore some points from A and B if the distance of the points exceeds ϵ.

2.3. Dimensional scaling with MDS

After acquiring the distance matrix using LCSS, dimension reduction is carried out to represent 3D data into 2D spatial data. MDS is a dimension reduction approach that preserves an object's core information while converting multidimensional data into a lower-dimensional space. The primary reason for utilizing MDS is to obtain a graphical representation of the data, making it easier to comprehend. There are some other dimensionality reduction techniques such as PCA, factor analysis, and isomaps. However, MDS is the most popular among these techniques due to its simplicity and various application areas [29]. MDS analysis to find spatial maps for objects is based on similarity or difference information between those objects.

2.4. Clustering with DBSCAN

Following the conversion of the distance matrix into spatial data, clustering is conducted with DBSCAN. The measurement of distance with DBSCAN spatial data can be done by calculating the Euclidean distance. Moreover, the DBSCAN algorithm is used to identify clusters and noise with the specified parameters *Eps* and minimum points (*MinPts)*. After completing the clustering process, the cluster labels will be visualized back to each trajectory. DBSCAN is also a density-based clustering algorithm, which scans for a high-density data set to serve as a cluster. DBSCAN does not estimate the density between points for efficiency reasons. Within a radius of the core point, all neighbors are regarded to be part of the same cluster as the core point [30]. The cluster shape generated by DBSCAN is density-dependent, and it is possible to generate arbitrary cluster shapes [31]. A cluster in DBSCAN is defined as the maximum data set connected within that density (density-connected). Membership of each profile is calculated based on the distance formula. Moreover, DBSCAN is considered an unsupervised clustering algorithm because the number of clusters generated is determined by the shape of the data distribution itself, not initialized at the beginning.

3. RESULTS AND DISCUSSION

This study uses datasets from terrestrial AIS receivers at Udayana University. The dataset used has 640,527 rows. Based on MMSI we found 437 vessels. The other attributes of the AIS data used in this study are timestamp, MMSI, latitude, longitude, SOG and COG. The experiment was carried out using M1 Macbook Air. Table 1 is details of the research instrument specifications.

Table 1. Research instrument

Item	Configuration
Number of rows	640,527
Number of vessel (MMSI)	437
AIS dataset	Udayana University terrestrial AIS receiver. Scope from latitude. -8.2 to longitude 116
Dataset stored at	MySQL 8 and .npy file format
Programing language	Python 3.8 with scikit-learn
Hardware spec.	8 Core Apple M1 CPU; 8GB LPDDR4X-4266 MHz SDRAM; 512GB NVMe SSD

3.1. Data preprocessing

There are three steps in the preprocessing stage. Figure 2 visually shows the change of data at each preprocessing step. The first step is the data cleaning, where abnormal data such as empty vessel position attributes, COG values outside 0-360, and out-of-range vessels positions are eliminated. In this study, we aim to identify the trajectory. Therefore, the data with a SOG value below 1.5 will also be eliminated because it shows a vessel is not moving.

The second step is to perform trajectory extraction. After the extraction, it is still necessary to perform data elimination for trajectories that only have a few rows. The data cleaning and trajectory extraction process succeeded in reducing the initial data in Figure 2 (a), which has 640,527 rows and 437 vessels, to Figure 2 (b), which has 127,144 rows and 231 vessels with 405 vessel trajectories.

The last step is to implement the DP algorithm to to compress each trajectory. The epsilon configuration used is 0.001, which is 111m. Figure 2 (c) shows that the number of rows of data can be reduced to 4,225. Visually, the shape of the trajectories maintains the same characteristics as the trajectories before compression. Table 2 shows the breakdown of data changes from the data preprocessing stages.

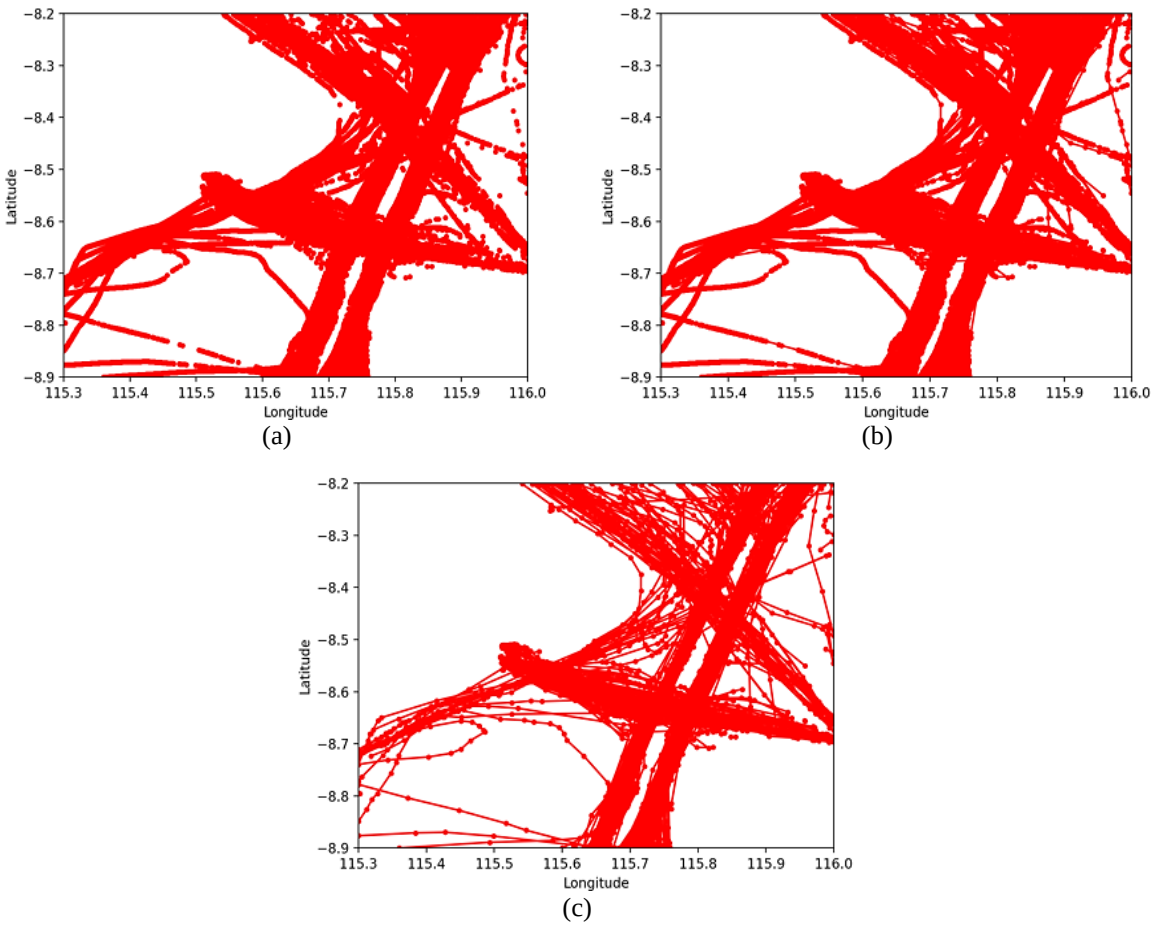


Figure 2. Data preprocessing step from (a) raw AIS data, (b) clean and extracted trajectories and (c) compressed trajectories

Table 2. Data preprocessing result

	RAW	Clean	Compressed
Number of rows	640,527	127,144	4,225
Number of vessel (MMSI)	437	231	231
Number of trajectories	-	405	405

3.2. Similarity measurement with LCSS

The LCSS algorithm is applied to measure the similarity distance between all trajectories in the similarity measurement stage with threshold parameter 0.1. The measured trajectories are trajectories that

have been compressed with the DP algorithm. The process to get the distance matrix took 19.414s. Figure 3 (a) is a 2D view of the distance matrix where the x-axis and y-axis are the vessel's trajectory. Figure 3 (a) shows the characteristics of the distance between trajectories. If the distance is close to 0, it is marked with a dark color indicating the similarity of the trajectory characteristics. On the other hand, if the value is greater than 0, it is marked with a light color to show differences between trajectories. In Figure 3 (b), the x-axis is the distance value, and the y-axis is the frequency of the number of passes. Figure 3 (b) also shows the number of similarities between the trajectories for each existing distance value.

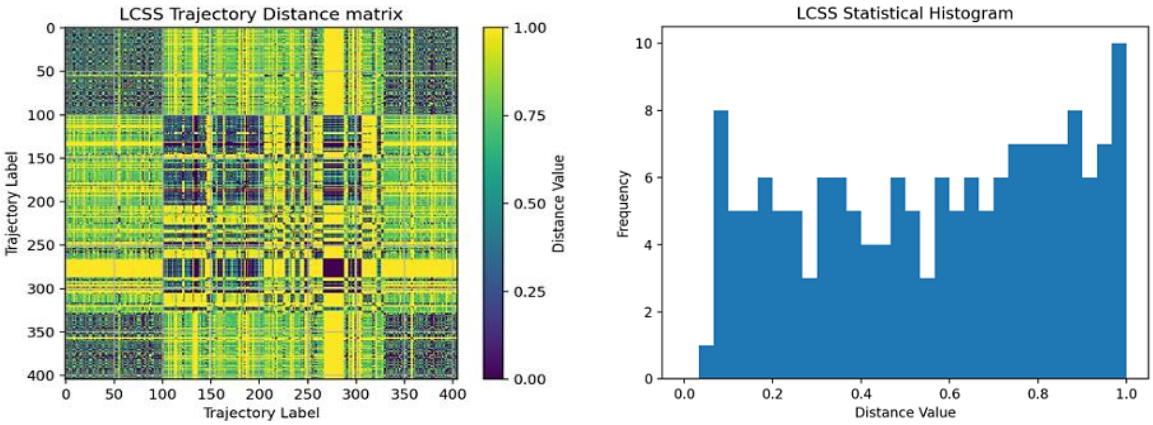


Figure 3. Similarity measurement between all trajectories (a) 2D image from distance matrix and (b) statistical histogram of all distance

3.3. Dimensional reduction with MDS

In the dimensional reduction process, the MDS algorithm converts the distance matrix from 3D data into 2D spatial data. The 2D distance matrix in Figure 3 (a) is 3D data where the x-axis and y-axis are trajectories labels. The z-axis is the value of the distance between trajectories. Figure 4 is the result of the MDS, which represents the data into 2D spatial data.

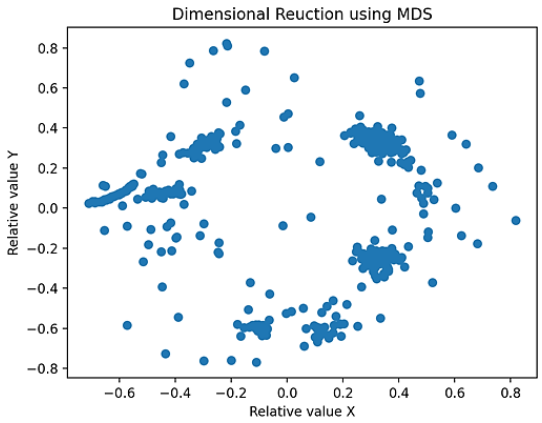
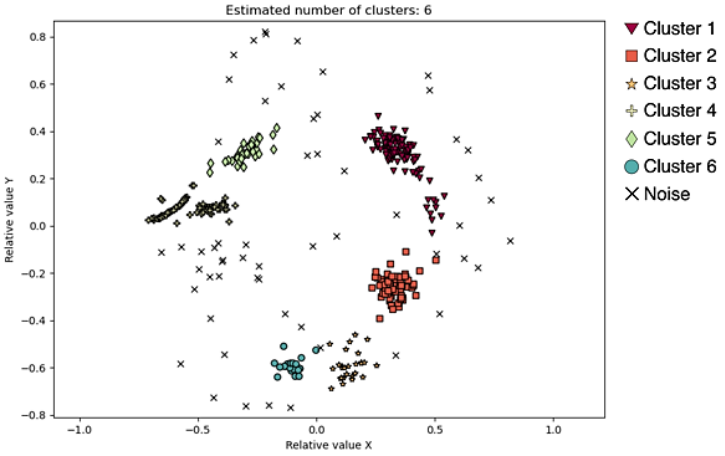


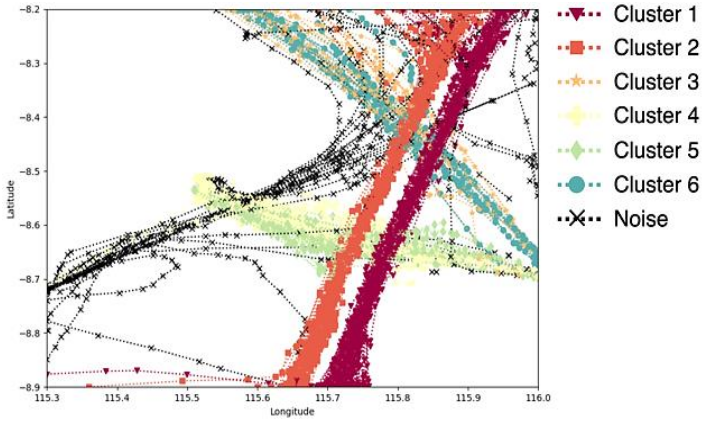
Figure 4. Dimensional reduction MDS spatial representation

3.4. Clustering with DBSCAN

The clustering stage uses the DBSCAN algorithm. The configuration used is *Eps*=0.088 and *MinPts*=9. The data exploited in the clustering process is the spatial data from the MDS in Figure 4, while the obtained result from clustering is shown in Figure 5 (a). The clustering results are then mapped to each trajectory, as shown in Figure 5 (b). Every color shows the trajectories cluster, except the colored black representing noise, where the black trajectories do not fit into any cluster.



(a)



(b)

Figure 5. Clustering result of (a) MDS representation and (b) trajectories labeled by cluster

3.5. Visualization of clustering result

The clustering results using the proposed clustering framework achieved an SC score of 0.533. Thus, the number of successfully generated clusters are 6, while 57 trajectories were found to be noise. The number of trajectories in each cluster can be seen in Table 3.

Each vessels trajectories clusters can be seen in figure 6. The clusters in Figures 6 (a) and 6 (b) show the trajectories that pass through the traffic separation scheme (TSS) in the Lombok Strait. Figure 6 (a) shows vessel traffic moving from south to north, and Figure 6 (b) shows the opposite direction. Figure 6 (c) is the vessel's trajectory from western Indonesia to Lombok. Figures 6 (d) and 6 (e) show the crossing routes that pass through the TSS on Lombok Strait. Figure 6 (d) illustrates the trajectory of vessels going from Lombok to Karangasem Bali, and Figure 6 (e) is for the opposite direction. Figure 6 (f) is the vessel's trajectory from Lombok to western Indonesia. Those figures indicate that the proposed LCSS clustering framework has succeeded in distinguishing trajectories that have different directions even though they have a similar trajectory shape visually.

Table 3. Trajectories in cluster result

No.	Cluster	Number of trajectories
a	1st Cluster	100
b	2nd Cluster	77
c	3rd Cluster	27
d	4th Cluster	88
e	5th Cluster	35
f	6th Cluster	21
g	noise	57

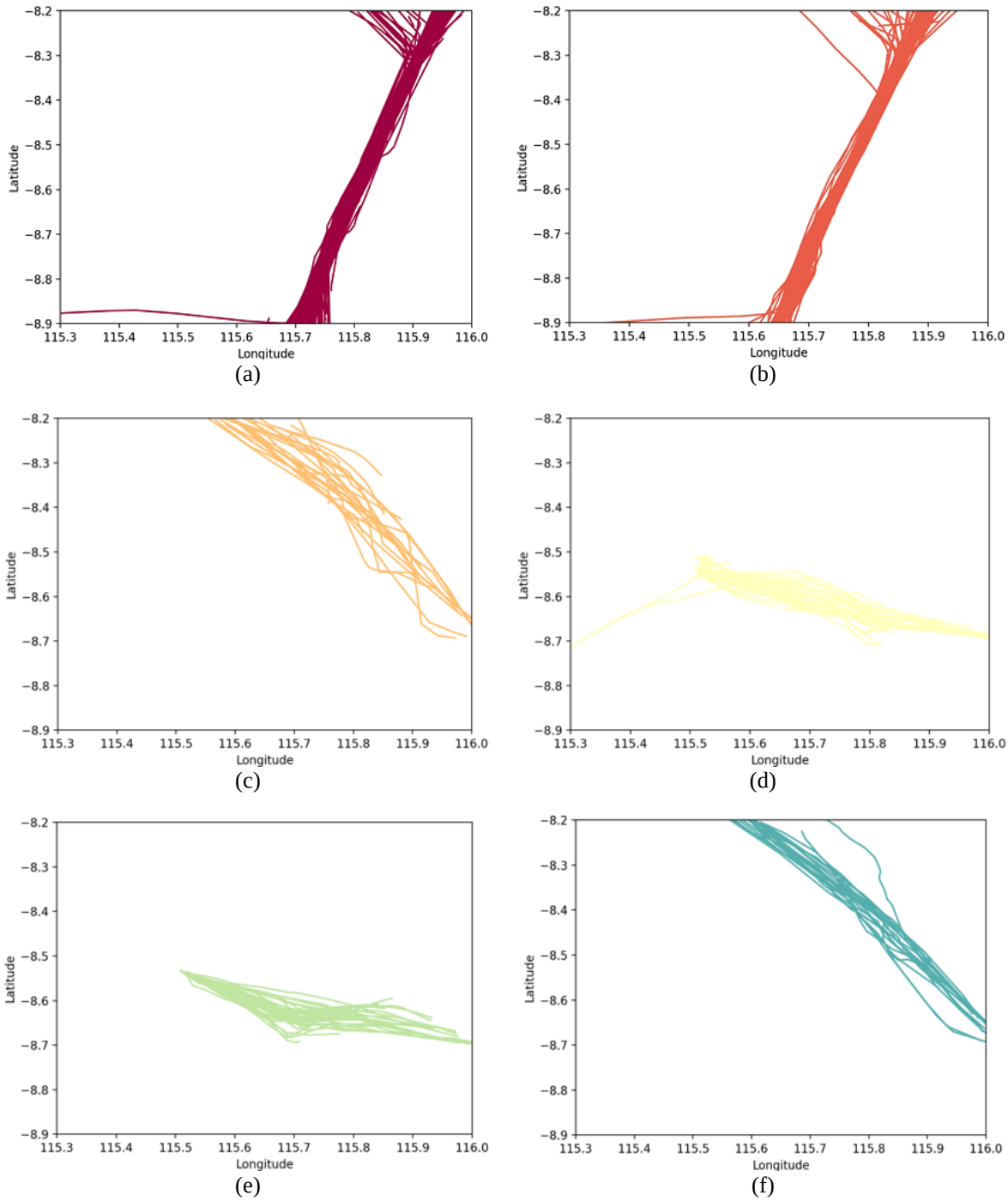


Figure 6. Vessel trajectory cluster of (a) Lombok Strait TSS south to north and (b) is the opposite; (c) Western Indonesia to Lombok; (d) Lombok to Karangasem and (e) is the opposite; and (f) Lombok to Western Indonesia

Figure 7 shows the visual trajectories that are included in the noise cluster. Trajectories that are included in the noise cluster are shorter in length in comparison to trajectories of vessels from southern Bali to northern Bali or vice versa. Those trajectories might be categorized as noise because the amount of data is very small.

3.6. Comparison with different algorithms

Here we provide the comparison of three similarity measurement algorithms, namely DTW, LCSS, and HD. Each algorithm uses the same compressed trajectory data. Three clustering algorithms were also compared, namely DBSCAN K-means and K-medoids. Table 4 shows the comparative description of each

method. As shown in Table 4, the clustering process using the HD algorithm is the fastest, with only 18,465s. However, the HD algorithm cannot distinguish trajectories in the opposite direction because it only measures the trajectory distance based on the shape of the trajectory. DTW takes the longest time with a total clustering time of 34,886s. DTW is very affected by abnormal AIS trajectory data, so it cannot provide optimal distance between trajectories. The results of clustering with DTW get the lowest SC score, which is 0.135. The similarity measurement algorithm that can distinguish the direction of the trajectory with a high SC score is LCSS with a total clustering time of 23,468s. The comparison of clustering algorithms is carried out using the results of the most optimal similarity matrix in the previous comparison, namely LCSS. The parameters used in each algorithm are the parameters with the highest SC score. The K-means and K-medoids algorithms cannot identify the noise. Both algorithms get a high SC score while recognizing 4 clusters. LCSS+DBSCAN is the only one that can recognize noise, getting 6 clusters with an SC score of 0.533. This comparison shows that the framework with the proposed combination of algorithms can solve the problem of similarity measurement to noisy AIS data. The proposed framework can also distinguish the direction of the trajectory with a relatively fast total clustering time.

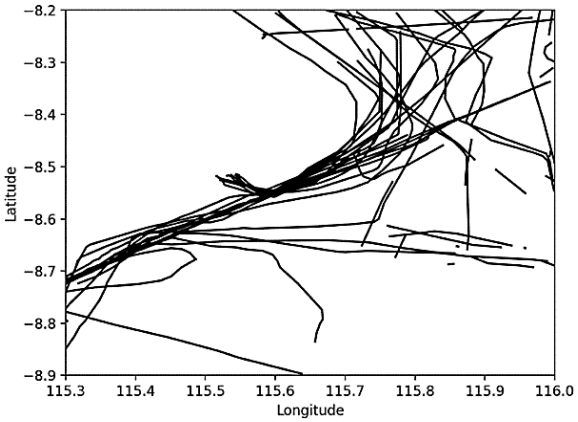


Figure 7. Trajectory noise

Table 4. Comparison results of different algorithms

No.	Proposed method LCSS+DBSCAN	DTW+DBSCAN [9]	HD+DBSCAN [16]	LCSS+K-means	LCSS+K-medoids
Opposite course	Yes	Yes	No	Yes	Yes
Detect noise	Yes	Yes	Yes	No	No
SC score	0.533	0.135	0.498	0.638	0.635
N cluster	6	6	6	4	4
Total time	23,468s	34,886s	18,465s	23,474s	23,472s

4. CONCLUSION

Trajectory clustering based on AIS data requires well-structured preprocessing steps due to the existence of some abnormal data. Moreover, the trajectory cannot be directly clustered with the clustering algorithm alone. Therefore, we propose a framework that combines several algorithms that can process AIS data from scratch to generate clusters. The main contribution of the proposed framework is a well-structured combination of algorithms in preprocessing, similarity measurement, and clustering to construct good quality clusters while minimizing total processing time. Our experiment shows that similarity measurement is the process that takes the longest time, and the chosen trajectory compression with DP significantly accelerates the process. We also observed that the LCSS algorithm is the optimal algorithm in similarity measurement of vessel trajectories based on AIS data. Furthermore, we found the right combination of MDS and DBSCAN for density-based clustering. The comparison in similarity measurement with DTW and HD, and the comparison of clustering with K-means and K-medoids show the performance advantage of the framework with the proposed combination of algorithms. Moreover, the proposed framework can distinguish trajectories in different directions, identify the noise, and produce clusters of good quality with relatively fast total processing time. However, the proposed framework still requires parameter determination for the DP, LCSS, and DBSCAN algorithms. Therefore, our future work will focus on investigating a parameter-free trajectory clustering framework for AIS data.

ACKNOWLEDGEMENTS

This research was supported and funded by the Ministry of Education, Culture, Research and Technology in the Republic of Indonesia with a contract number: B/136-18/UN14.4.A/PT.01.05/2021.

REFERENCES

[1] L. Li, W. Lu, J. Niu, J. Liu, and D. Liu, "AIS data-based decision model for navigation risk in sea areas," *Journal of Navigation*, vol. 71, no. 3, pp. 664–678, May 2018, doi: 10.1017/S0373463317000807.

[2] Z. Yan *et al.*, "Exploring AIS data for intelligent maritime routes extraction," *Applied Ocean Research*, vol. 101, pp. 1–10, Aug. 2020, doi: 10.1016/j.apor.2020.102271.

[3] X. Wu, A. Rahman, and V. A. Zaloom, "Study of travel behavior of vessels in narrow waterways using AIS data – A case study in Sabine-Neches Waterways," *Ocean Engineering*, vol. 147, pp. 399–413, Jan. 2018, doi: 10.1016/j.oceaneng.2017.10.049.

[4] K. T. Seong and G. H. Kim, "Implementation of voyage data recording device using a digital forensics-based hash algorithm," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6, pp. 5412–5419, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5412-5419.

[5] S. Arasteh *et al.*, "Fishing vessels activity detection from longitudinal AIS data," in *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, Nov. 2020, pp. 347–356. doi: 10.1145/3397536.3422267.

[6] M. Fournier, R. C. Hilliard, S. Rezaee, and R. Pelot, "Past, present, and future of the satellite-based automatic identification system: Areas of applications (2004–2016)," *WMU Journal of Maritime Affairs*, vol. 17, no. 3, pp. 311–345, Sep. 2018, doi: 10.1007/s13437-018-0151-6.

[7] M. Svanberg, V. Santén, A. Hörteborn, H. Holm, and C. Finnsgård, "AIS in maritime research," *Marine Policy*, vol. 106, pp. 1–10, Aug. 2019, doi: 10.1016/j.marpol.2019.103520.

[8] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1559–1582, May 2018, doi: 10.1109/TITS.2017.2724551.

[9] H. Li, J. Liu, R. Liu, N. Xiong, K. Wu, and T. Kim, "A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis," *Sensors*, vol. 17, no. 8, p. 1792, Aug. 2017, doi: 10.3390/s17081792.

[10] H. Li, J. Liu, K. Wu, Z. Yang, R. W. Liu, and N. Xiong, "Spatio-temporal vessel trajectory clustering based on data mapping and density," *IEEE Access*, vol. 6, pp. 58939–58954, 2018, doi: 10.1109/ACCESS.2018.2866364.

[11] P. Sheng and J. Yin, "Extracting shipping route patterns by trajectory clustering model based on automatic identification system data," *Sustainability*, vol. 10, no. 7, pp. 1–13, Jul. 2018, doi: 10.3390/su10072327.

[12] P. Last, C. Bahlke, M. Hering-Bertram, and L. Linsen, "Comprehensive analysis of automatic identification system (AIS) data in regard to vessel movement prediction," *Journal of Navigation*, vol. 67, no. 5, pp. 791–809, Sep. 2014, doi: 10.1017/S0373463314000253.

[13] I. P. N. Hartawan, I. M. O. Widyantara, A. A. I. N. E. Karyawati, N. I. Er, K. B. Artana, and N. P. Sastra, "AIS data pre-processing for trajectory clustering data preparation," in *2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES)*, Nov. 2021, pp. 1–5. doi: 10.1109/ICARES53960.2021.9665187.

[14] E. Tu, G. Zhang, S. Mao, L. Rachmawati, and G.-B. Huang, "Modeling historical AIS data for vessel path prediction: A comprehensive treatment," pp. 1–11, Jan. 2020, [Online]. Available: http://arxiv.org/abs/2001.01592

[15] Y. Zhang and G. Shi, "Trajectory similarity measure design for ship trajectory clustering," in *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*, Mar. 2021, pp. 181–187. doi: 10.1109/ICBDA51983.2021.9403137.

[16] W. Yitao, Y. Lei, and S. Xin, "Route mining from satellite-AIS data using density-based clustering algorithm," *Journal of Physics: Conference Series*, vol. 1616, no. 1, pp. 1–8, Aug. 2020, doi: 10.1088/1742-6596/1616/1/012017.

[17] R. Zhen, Y. Jin, Q. Hu, Z. Shao, and N. Nikitakos, "Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naïve bayes classifier," *Journal of Navigation*, vol. 70, no. 3, pp. 648–670, May 2017, doi: 10.1017/S0373463316000850.

[18] L. Wang, P. Chen, L. Chen, and J. Mou, "Ship ais trajectory clustering: An hdbscan-based approach," *Journal of Marine Science and Engineering*, vol. 9, no. 6, pp. 1–20, May 2021, doi: 10.3390/jmse9060566.

[19] L. Zhao and G. Shi, "A novel similarity measure for clustering vessel trajectories based on dynamic time warping," *Journal of Navigation*, vol. 72, no. 2, pp. 290–306, Mar. 2019, doi: 10.1017/S0373463318000723.

[20] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 123–144, Jan. 2017, doi: 10.1007/s10462-016-9477-7.

[21] Z. Chen, J. Guo, and Q. Liu, "DBSCAN algorithm clustering for massive AIS data based on the hadoop platform," in *2017 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*, Dec. 2017, pp. 25–28. doi: 10.1109/ICIICII.2017.72.

[22] M. Z. Hossain, M. J. Islam, M. W. R. Miah, J. H. Rony, and M. Begum, "Develop a dynamic DBSCAN algorithm for solving initial parameter selection problem of the DBSCAN algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1602–1610, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1602-1610.

[23] Y. Chen, S. Tang, N. Bouguila, C. Wang, J. Du, and H. Li, "A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data," *Pattern Recognition*, vol. 83, pp. 375–387, Nov. 2018, doi: 10.1016/j.patcog.2018.05.030.

[24] A. Starczewski and A. Cader, "Determining the eps parameter of the DBSCAN algorithm," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, pp. 420–430. doi: 10.1007/978-3-030-20915-5_38.

[25] U. N. Wisesty and T. R. Mengko, "Comparison of dimensionality reduction and clustering methods for SARS-CoV-2 genome," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2170–2180, Aug. 2021, doi: 10.11591/eei.v10i4.2803.

[26] G.-Y. Zhang and J. Zhang, "Trajectory clustering based on trajectory structure and longest common subsequence," in *International Conference on Computer, Electronic Information and Communications (CEIC 2018)*, Aug. 2018, pp. 61–65. doi: 10.12783/dtcse/ceic2018/24525.

[27] A. Octavian, T. Trismadi, and P. Lestari, "The importance of establishing particularly sensitive sea areas in Lombok Strait: Maritime security perspective," *IOP Conference Series: Earth and Environmental Science*, vol. 557, no. 1, pp. 1–12, Aug. 2020, doi: 10.1088/1755-1315/557/1/012013.

[28] L. Zhao and G. Shi, "A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm," *Ocean Engineering*, vol. 166, pp. 37–46, Oct. 2018, doi: 10.1016/j.oceaneng.2018.08.005.
[29] N. Saeed, H. Nam, M. I. U. Haq, and D. B. M. Saqib, "A survey on multidimensional scaling," *ACM Computing Surveys*, vol. 51, no. 3, pp. 1–25, May 208AD, doi: 10.1145/3178155.
[30] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 1–21, Aug. 2017, doi: 10.1145/3068335.
[31] Y. H. R. *et al.*, "Massively scalable density based clustering (DBSCAN) on the HPCC systems big data platform," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 207–214, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp207-214.

BIOGRAPHIES OF AUTHORS

I Made Oka Widyantara received the B.Eng. in electrical engineering from Sepuluh Nopember Institute of Technology (ITS) Surabaya, Indonesia, in 1997. He completed M. Eng degree in the Telecommunication Information System from Bandung Institute of Technology (ITB), Bandung, Indonesia, 2001 and Dr. degree in the multimedia telecommunication from Sepuluh Nopember Institute of Technology (ITS) Surabaya, Indonesia, 2013, He is currently a lecturer in the Department of Electrical Engineering with Udayana University and head of Information Technology Department, Engineering Faculty, Udayana University, Indonesia. His research interests include intelligent signal processing, and image/video compression. He can be contacted at email: oka.widyantara@unud.ac.id.

I Putu Noven Hartawan received the B. Eng degree in Informatics Engineering from STMIK STIKOM Indonesia in 2018. Currently he is pursuing his M. Eng degree at the Postgraduate Program in Electrical Engineering, Udayana University. His interest is research on data mining and software architecture and optimization. He can be contacted at email: novenhartawan@gmail.com.

Anak Agung Istri Ngurah Eka Karyawati received the M. Eng degree in Information Science and Systems Engineering from Ritsumeikan University, Japan, and Dr. degree in Doctoral Program of Computer Sciences, Gadjah Mada University, Indonesia. She is currently a researcher and a lecturer in Informatics Department, Udayana University, Bali, Indonesia. Her research interests include text mining, computational language, natural language processing, information retrieval, and knowledge representation. She can be contacted at email: eka.karyawati@unud.ac.id.

Ngurah Indra Er is a lecturer in the Department of Electrical Engineering of the Udayana University (UNUD), Bali, Indonesia, since 2002. He obtained his PhD from IMT Atlantique, Rennes, France in 2021. He completed his M.Sc. in communications engineering from University of Birmingham, U.K., in 1999 after receiving his B.Eng. in electrical engineering from Sepuluh Nopember Institute of Technology (ITS) Surabaya, Indonesia, a year earlier. His current research interests include Internet of Things (IoT), Internet of Vehicles (IoV), vehicular networks, opportunistic networks, and smart city data collection. He can be contacted at email: indra@unud.ac.id.

Ketut Buda Artana received the M. Sc degree in Marine Engineering from the University of Newcastle Upon Tyne–UK in 1997 and Dr degree in the same field from Kobe University of Mercantile Marine, Japan in 2003, He is currently a professor in Maritime Reliability and Safety, Department of Marine Engineering, Faculty of Marine Technology ITS Surabaya. With colleagues, he has been developing and commercializing AISITS, a real time early warning system for protecting ships, subsea pipeline, subsea cable, and offshore platform. His current position is the head of safety laboratory and senior researcher at the Center of Excellence of Maritime Safety and Marine Installation (PUI-KEKAL). His research interests include maritime risk management, marine safety assessment, reliability engineering, LNG technology. He can be contacted at email: ketutbuda@its.ac.id.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 12∼22
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp12-22

Information-gathering dialog system using acoustic features and user's motivation

Ryota Togai, Takashi Tsunakawa, Masafumi Nishida, Masafumi Nishimura
Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University, Hamamatsu, Japan

Article Info

Article history:

Received Nov 14, 2021
Revised Sep 3, 2022
Accepted Sep 22, 2022

Keywords:

Chat dialog system
Information-gathering dialog system
Nonverbal acoustic features
Talking motivation
Topic induction

ABSTRACT

Recently, as society continues to age, automation of watching elderly people who live apart from their families has been gradually expected. However, we must prevent them from losing their purpose in life, which declines due to lack of communication. Thus, a chat dialog system has attracted widespread attention as a method that achieves both problems: keeping their purpose in life and watching their daily lives. Unlike a task-oriented dialog system, a chat dialog system has explicitly no task to accomplish and makes a conversation to continue communication with the users. Keeping a conversation is essential for elderly people who are mostly unfamiliar with digital devices. Moreover, conversing daily on the chat dialog system provides the opportunity to collect information for their care. This study realizes an information-gathering dialog system, a chat dialog system that collects healthcare information of elderly people. Furthermore, we use the nonverbal acoustic features from their speech, since automatic speech recognition is not necessarily accurate in current systems. This paper illustrates the effectiveness of two important elements, topic change for keeping the talking user motivated with the dialog system and motivation estimation, for attaining an information-gathering dialog system using nonverbal acoustic features.

Corresponding Author:

Takashi Tsunakawa
Department of Computer Science, Faculty of Informatics, Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka 432-8011 Japan
Email: tuna@inf.shizuoka.ac.jp

1. INTRODUCTION

Recently, society of many developed countries has noticed an increased aging population, leading to the automation of watching elderly people who live apart from their families [1], [2]. These elderly people living away from their families lack communication and eventually lose something to live for. To solve this problem, a chat dialog system has attracted widespread attention as a method for keeping their purpose in life and watching their daily lives [3], [4]. By having the elderly people talk with the chat dialog system daily, the system helps increase chances of communication and obtain the necessary information for monitoring the elderly people by asking important and timely questions. Studies over the years have reported methods for gathering information from users through dialog systems. In information-gathering dialog systems, Kobayashi *et al.* [3] highlighted the difficulty for users to answer questions if the dialog system sequentially asks questions to be answered without considering the context of the dialog. They proposed a method using the chain structure of dialogs, which gradually shifts dialog topics to follow the dialog context and ask the questions to be answered. Such a topic shifting is called topic induction. Nagasaka *et al.* [5] used WordNet to build a topic induction model that shifts a current topic to the specified one in chat dialogs, to automate questions on

dementia in chat dialogs with elderly people. Yoshito *et al.* [6] aimed to build an active information-gathering dialog system, and created a model that determines the user's intention to end the dialog from nonverbal acoustic information to avoid the system asking persistent questions. Ishihara *et al.* [7] performed the interviewee's dialog willingness estimation to calculate questioning strategies in real-time for an information-gathering dialog robot. Previously, studies have considered dialog management, including topic induction algorithms, using features based mainly on linguistic information. Meguro *et al.* [8] have employed partially observable markov decision process (POMDP), which is a statistical dialog management method that sets rewards for actions in a probabilistically determined state transition structure, and executes actions that maximize the rewards that can be obtained in the future. Lison [9] have developed a dialog system that combines statistical dialog management and rule-based dialog management, which is available as an open source software [10]. However, it has not been sufficiently studied on appropriate timing and destination of topic induction considering nonverbal acoustic information.

However, simply talking without any consideration does not automate monitoring. For example, asking many questions to collect a lot of information ends up like a questionnaire session with a chat dialog system, which would not be appreciated by the elderly. On the other hand, by pursuing only the naturalness of the dialog, the system fails to ask the questions required of a monitoring chat dialog system. Additionally, one major challenge is topic transition. In information-gathering dialogs, the problem is reflected in how to efficiently move from the current topic to a new topic for the system to talk about. Humans recognize mental distances between topics in conversations, and feel uncomfortable when conversation suddenly moves to a distant topic or stays on near topics for so long. To solve these problems, estimating the user's talking motivation on the current topic is essential. By understanding the user's talking motivation, the system decides when to ask appropriate questions and which topic to move to. To make users continue talking with dialog systems daily, appropriately switching topics to talk is necessary. To achieve this, the dialog system judges whether it changes the current topic according to user's talking motivation or topic interest. Yokoyama *et al.* [11] developed a chat dialog system that switches the system's role to "listener" and "speaker" depending on the user's interest.

Previous studies on estimating users' talking motivation have used facial images, voice, and linguistic information of the user. Schuller *et al.* [12] studied to estimate the user's interest in current topics from multimodal information of facial expressions, nonverbal acoustic information, and verbal information obtained from a single speech of the user. Chiba *et al.* [13] automatically estimated talking motivation from multimodal information to build an interview dialog system. Saito *et al.* [14] estimated users' attitudes toward dialog from multimodal information in dialog data with dementia patients. Many previous studies have estimated the user's talking motivation using multimodal information. However, when one uses the dialog system, simultaneously capturing the user's facial expressions with cameras or performing complete speech recognition to acquire linguistic information is difficult. Since the dialog state changes gradually through multiple turns, efficiently learning the information of multiple turns is necessary. In dialog-state tracking challenge (DSTC) [15], a shared task that analyzes dialog using information from multiple turns, methods using recurrent neural network (RNN) has shown high performance [16]. In these methods, using the linguistic information of the user's speech as input, the probability distribution of tasks, user's requests, and so on are estimated as dialog states. A dialog-state tracking method using long short term memory (LSTM), which improves the drawback of RNNs with difficulty storing long-term information, has been proposed [17]. We consider that the dialog-state tracking is very similar to the task of measuring user's talking motivation, since the motivation can be regarded as a kind of dialog states. We apply this dialog-state tracking method to track the user's talking motivation using RNN with nonverbal acoustic information as the user input.

In this study, we experiment by measuring the degree of user satisfaction when the Wizard of Oz system [18], [19] switches topics according to the user's estimated talking motivation with the current topic. In addition, we focus on introducing nonverbal acoustic information for estimating the talking motivation. In human-human dialog, various nonverbal information such as prosody and facial expressions is also frequently used. Hence, such information has been considered important as an input to the dialog system [20], [21]. We analyze the relationship between nonverbal acoustic information and the talking motivation to be estimated.

2. THE PROPOSED METHOD

To collect information from users by asking questions during a chat dialog, question timing and topic transition must be adjusted appropriately. This section proposes two hypotheses about topic induction from the

topic space model. This section also verifies them using the Wizard of Oz.

2.1. Modeling of the topic space

We describe the topic space model proposed for realizing a dialog system with topic induction. Suppose the dialog system suddenly switches from the current topic to a mentally distant topic, the user will feel that the system skips from topic to topic. However, if the system repeatedly talks about similar topics, the user gets bored and the satisfaction of the dialog decreases. For a user to enjoy a chatting dialog system for a long time, the system must switch to distant or near topics at the right time. Therefore, it is important to model the topic space representing the mental distance among topics.

We model the topic space with a two-dimensional undirected graph structure reflecting mental distance among topics referring to Nagasaka *et al.*'s work [5]. Figure 1 shows an example of modeling the topic space. Each node represents a topic, and topics connected by an edge are mutually transitable. The length of an edge represents the mental distance between the topics. For users to feel naturally induced by these topics, gradually moving from the current topic to the goal topic in the topic space is essential.

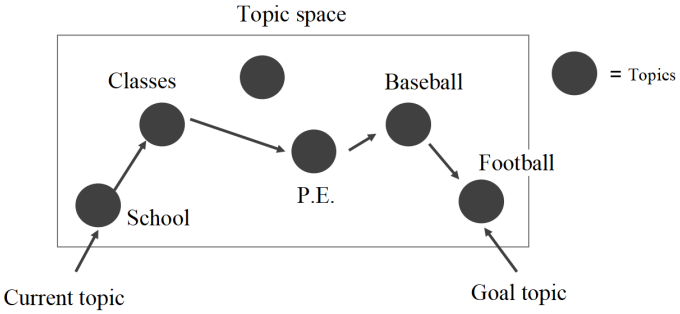


Figure 1. Example of modeling a topic space

Here, we model the topic space using WordNet [22] and Word2vec [23]. WordNet is a tree representation of the conceptual structure of words. Furthermore, the WordNet-based distance between concepts can be obtained by following the shortest path between nodes in the tree. suppose the distance between concepts is roughly consistent with the human mental scale. Then, we can apply this to the topic space model. Word2vec is a model representing words as vectors and obtains the similarity between two words by calculating the cosine of their vectors. The cosine similarity is negatively correlated with the mental distance between two words and can be used for a topic space. The Word2vec-based distance is the value of subtracting the Word2vec similarity between the keywords that indicate the topic from 1. We used Japanese Wikipedia as a training corpus to obtain word vectors by Word2vec.

2.2. Topic induction and user satisfaction

The straightforward way to obtain the required information is to directly ask questions about the information. However, as Kobayashi *et al.* [3] stated, this reduces user satisfaction. Therefore, to simultaneously maximize both user satisfaction and the amount of information obtained by the dialog system, we find the time at which user satisfaction does not decrease even if the topic is changed to ones the user asks a question about once. We formulated the following hypothesis about topic induction, referring to human interaction.

Hypothesis 1 *When the user's talking motivation with the current topic is low, switching to a distant topic does not decrease the user's satisfaction.*

In human-human conversation, if the person we talk to seems to enjoy the current topic, we delve deeper into the topic, otherwise, we change the topic to a different one to explore the person's talking motivation. If the same dialog strategy can be used for dialog systems, it would be possible to continue dialog without lowering user motivation by choosing topics close in the conceptual distance when the user's motivation for dialog is high and switching to farther topics otherwise. Also, we consider another hypothesis:

Hypothesis 2 *The user's talking motivation is correlated with features of nonverbal acoustic information, such as loudness and length of the user's speech.*

This is also supported by human-human conversation; loud voice and/or long speech of the person can indicate more motivation to talk about the current topic, whereas smaller voice and/or shorter replies show little motivation. Also, we estimate the user's talking motivation from features of nonverbal acoustic information.

3. METHOD

Our proposed model is based on two hypotheses described in the previous section. Here, we verify these hypotheses and the effectiveness of our topic induction strategy for collecting information by two experiments. One is estimating user's talking motivation, and the other is topic switching Wizard of Oz experiment for analyzing the talking motivation, topic distance, and user satisfaction.

3.1. Experiment 1: estimating user's talking motivation

First, we focus on showing the appropriateness of the hypothesis 2. To analyze and estimate users' talking motivation, we collected spoken dialog data with recorded talking motivation at each turn. The user talks with the dialog system through the microphone of the smartphone. The voice during the dialog is recorded using the microphone of a smartphone with a sampling frequency of 16 kHz and a quantization bit of 16 bits. Following the previous study [3], the system talks only one topic in one session and takes 20 turns as either a listener or a speaker. The system as a listener only asks questions to the user, and the system as a speaker only discloses itself to the user. The system employed the use of fixed scenarios based on fixed topics for speech, and no questions from the user were allowed. The user records his/her current talking motivation on a 7-point scale from -3 to 3 for each turn during the dialog. The first turn of the dialog is set to 3 because we assume that the user actively begins to talk with the dialog system. Five-topic scenarios were prepared for the system to talk about including computers, cooking, fashion, travel, and music. This follows the literature in [13] so that the level of users' interests would be distributed. The change in talking motivation depends on the level of interest in the topic. Therefore, the user's level of interest was recorded in each topic on a 5-point scale from -2 to 2 upon completing the dialog. To conduct each session independently, one session was held per day, and six subjects were asked to talk with the dialog system at home for ten days. From this experiment, audio data were obtained from 60 sessions with six subjects acting as listeners and speakers, respectively, for five topics.

3.2. Experiment 2: talking motivation and user satisfaction

To analyze the relationship between the user's talking motivation and the conceptual distance between topics, we conducted a Wizard of Oz dialog experiment with the topic switched according to the user's motivation for dialog. During the dialog, subjects inputted their motivation to talk about the current topic at each turn of the dialog in 11 levels: 0, 10, 20, ..., 100. The greater value showed higher motivation. The Wizard switched the topic every four turns according to users' talking motivation. Thus, for the two dialog sessions, each with a 10 min duration, the experiment for each of the 10 subjects is as:

- Session A: a session in which the system chooses a distant topic when the user's talking interest is 50 or more, and a closer topic otherwise.
- Session B :a session in which the system chooses a closer topic when the user's talking interest is 50 or more, and a distant topic otherwise.

The distance between topics is measured using the Wizard's mental scale. After the dialog, the subjects rated their satisfaction on a 7-point scale from -3 to 3. A higher value showed a higher level of satisfaction.

4. RESULTS AND DISCUSSION

4.1. Experiment 1: estimating user's talking motivation

4.1.1. User's interest in the topic

We analyze the effects of "user's interest in the topic" and "the role of the system as a listener or a speaker" among the factors considered influencing the user's talking motivation. Figure 2 shows a scatter plot of the slope of the change in user's talking motivation and the user's level of interest in the topic. Here, the slope of the change in the user's talking motivation is obtained from the slope of the linear regression calculated for the series of user's talking motivation for 20 turns. The distribution of the plots in the scatter plot is right-shouldered, and the slope of the regression line is positive, indicating that higher level of user's interest in the current topic positively affects, increases, or keeping user's talking motivation.

4.1.2. Role of the dialog system

Next, we analyzed the transition of a user's talking motivation depending on whether the system plays the role of a listener or a speaker. Figure 3 shows the average slope of the change in user's talking motivation for each user and system role. The error bars represent the standard deviation. From the figure, the slope of the change in users' talking motivation is negative for nearly all users, indicating that their talking motivation decreased as the dialog progressed. Thus, the role of the system as a listener or a talker had no significant effect on the user's talking motivation.

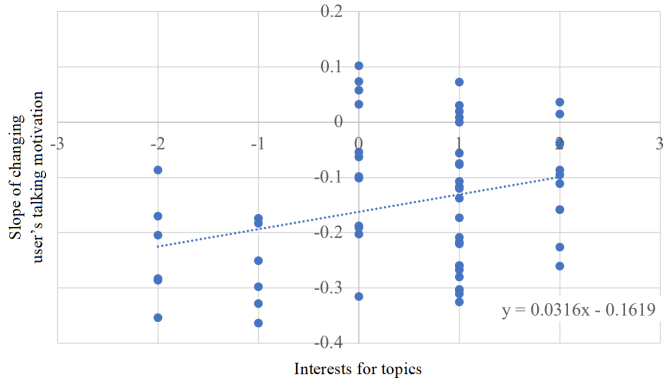


Figure 2. Scatter plot of interests for topics and slope of changing user's talking motivation

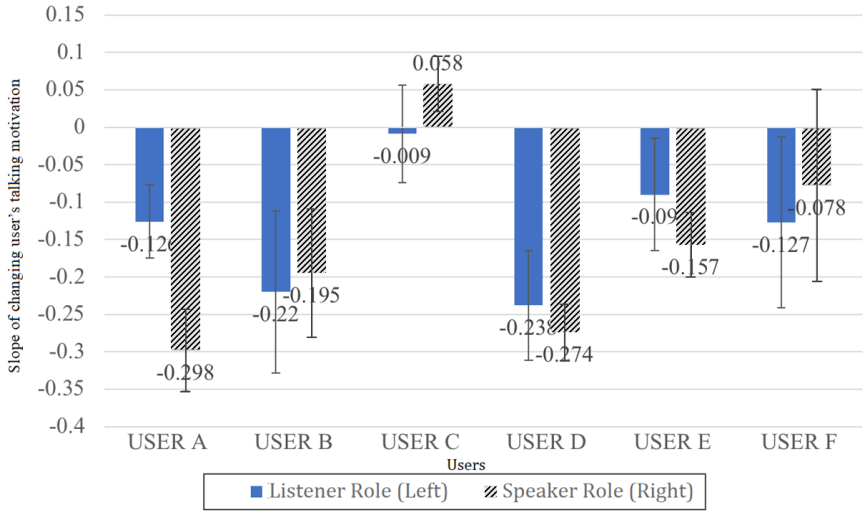


Figure 3. System roles and slope of changing user's motivation of each user

4.1.3. Nonverbal acoustic information

Estimating the user's talking motivation from factors, such as the role of the dialog system is difficult. Even by collecting each user's interesting topics, it is still difficult to estimate the current user's talking motivation due to the low correlation in Figure 2. To more directly estimate the user's talking motivation, we employ nonverbal acoustic information obtained from the user's speech. First, we deleted the silence before and after each turn of the audio data obtained from the user's dialog. Next, we extracted 384 features that can be extracted using openSMILE [24] IS09 emotion challenge configuration [25], which adds features of speech length, articulation rate, and delay. The delay feature encompasses the time from the end of the system speech to the beginning of user speech.

The nonverbal acoustic information extracted from the user speech during a dialog is highly dependent on the content of the speech, which significantly changes in a single turn. However, since the user's talking motivation does not change significantly from one turn to the next, the features for estimating the user's talking motivation become values that gradually change. Therefore, the extracted nonverbal acoustic information was smoothed by taking a five-point moving average in the turn direction for each session. This implies that the feature value of a given turn was the average of five turns, including both turns before and after the corresponding nonverbal acoustic information.

Also, we analyzed the correlation between the extracted nonverbal acoustic information and the user's talking motivation to investigate which nonverbal acoustic information is effective for the estimation [26]. Table 1 shows the top 10 features in the absolute value of the correlation coefficient. Among the nonverbal acoustic information, the correlation coefficient for the most strongly correlated feature was 0.311. No acoustic feature with a strong correlation was applied to all users. Furthermore, results showed that many mel-frequency cepstral coefficients (MFCC) ranges appeared in the top 10 features. Since the correlation coefficient is positive, the range of MFCCs became smaller as the user's talking motivation decreased.

Table 1. Top 10 features correlating with user's motivation

Feature	Correlation coefficient
(cf.) Turn number in session	-0.628
Voice rate	0.311
MFCC 8-dim. stddev	0.277
MFCC 8-dim. linregQ	0.260
Prob. of voice amean	0.257
Volume amean	0.248
MFCC 6-dim. range	0.236
MFCC 1-dim. range	0.227
MFCC 9-dim. stddev	0.227
MFCC 9-dim. linregQ	0.223

Figure 4 shows the maximum absolute value of the correlation coefficient calculated for each user. The maximum correlation coefficient exceeded 0.5 for many users, indicating that a correlation between nonverbal acoustic information and users' talking motivation exists. This result shows that there are individual differences in the relationship between nonverbal acoustic information and the user's talking motivation. Furthermore, it indicates that we can create a model with high accuracy by creating an individual estimation model for each user.

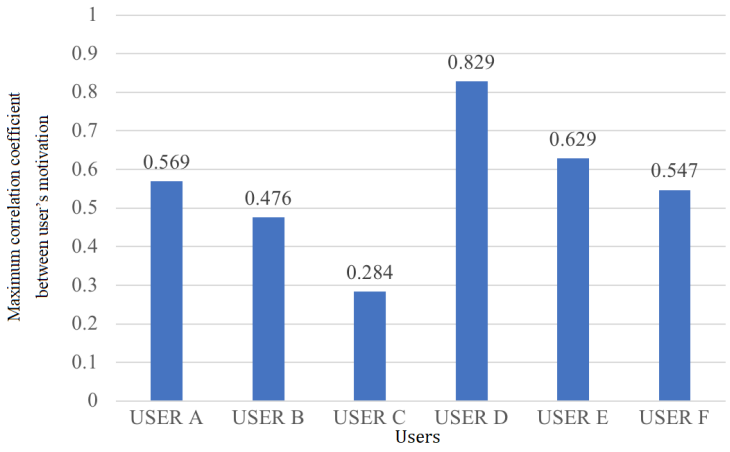


Figure 4. Maximum correlation coefficient between user's motivation for each user and acoustic features

4.1.4. Estimating user's talking motivation

This section compares the following three methods for estimating dialog motivation using nonverbal acoustic information of multiple turns:

– NN1 :a neural network (NN) that performs estimation using only features from one turn.
– NN3 :a NN that performs estimation using information from three previous turns.
– LSTM3 :an LSTM with a window size of three previous turns.

Each of them is evaluated using the mean absolute error (MAE) with a 10-point cross-validation.

Figure 5 shows the estimation accuracy of NN when only turn information is used as features and when nonverbal acoustic features are combined. The nonverbal acoustic features used were those of the top 20 correlations with users' talking motivation. The estimation error with nonverbal acoustic information was 0.451 lesser in MAE than that without nonverbal acoustic information, indicating that nonverbal acoustic information is an effective feature in estimating the user's talking motivation.

Figure 6 compares the error in estimating the user's talking motivation among the three estimation methods (here, the turn information not included in the nonverbal acoustic information is not used). The blue and orange bars show the results for the top 20 and top 300 correlated features, respectively. From Figure 6, for both the top 20 and top 300 features, the estimation error for using multiple turns of information was smaller than using only one turn of information. Also, the error was smallest when using LSTM. This indicates that the information from multiple turns is effective for the estimation and that LSTM reduces the estimation error.

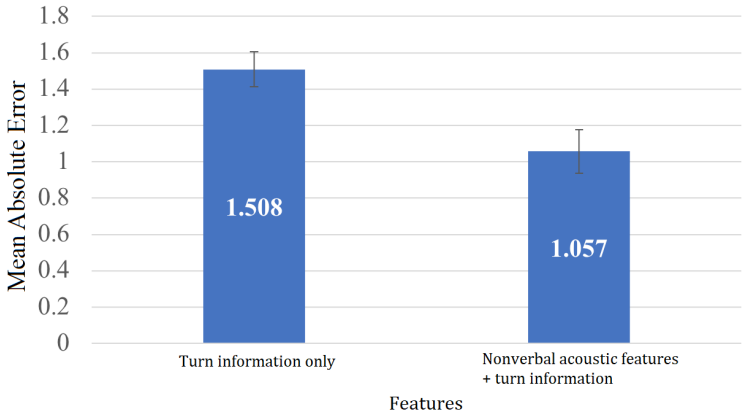


Figure 5. Comparison of estimation errors with and without nonverbal acoustic features

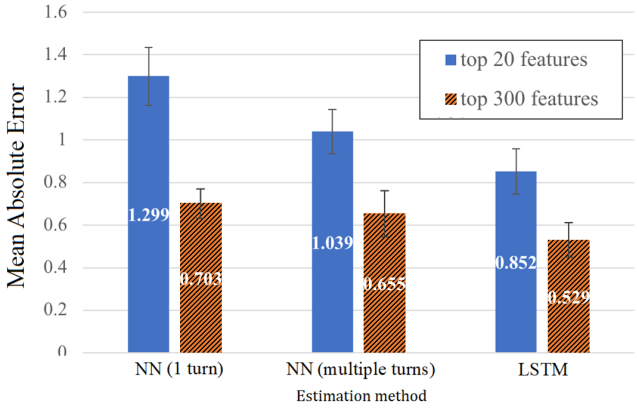


Figure 6. Comparison of estimation errors among estimation methods

4.2. Experiment 2: talking motivation and user satisfaction

4.2.1. Topic distance

This section checks whether the wizard chooses between distant and near topics according to the human mental scale. The relationship between the subject's talking motivation at the timing of topic switching and the conceptual distance between the previous and following topics is shown in Figure 7. In session A, the higher the subject's talking motivation, the more distant the wizard chose the topic, thereby creating a right-shouldered regression line. However, session B has the opposite strategy and has seen a steady increase. Therefore, sessions A and B have data that conformed to the conditions for topic selection, as shown in hypothesis 1. However, the slope of the regression line is not large, confirming the gap between the conceptual distance of WordNet and the human mental scale.

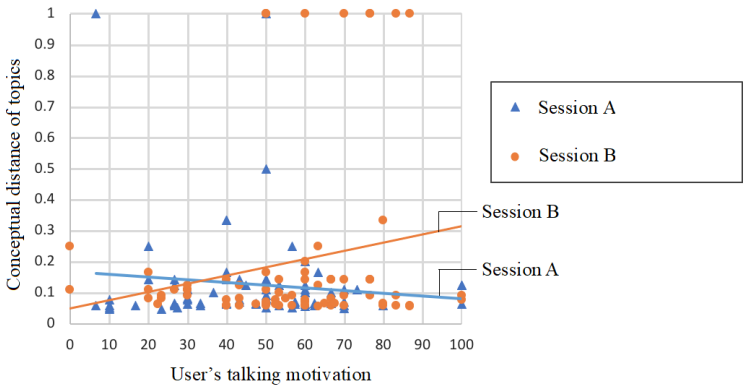


Figure 7. Correlation between user's talking motivation and conceptual distance of topics (calculated using WordNet)

Figure 8 shows the scatterplot relationship between the talking motivation and conceptual distance between topics. Here, the conceptual distance is calculated using Word2vec trained from Japanese Wikipedia. The results showed that the slope was larger than that of Figure 7 and that the conceptual distance when modeling the topic space can be modeled closer to the human mental scale using Word2vec.

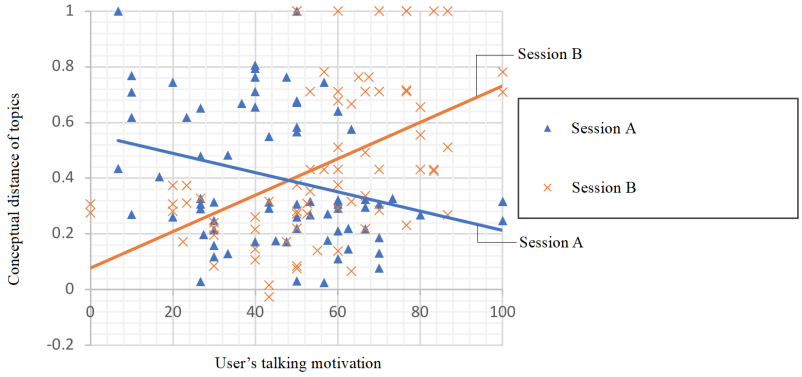


Figure 8. Correlation between users' motivation and concept distance of topics (calculated using Word2vec)

4.2.2. User satisfaction

The results of user satisfaction are shown in Figure 9. The average score was 1.0 higher in Session B than in Session A. Despite variations in the scale for each user's satisfaction, the results for each user show that most users were more satisfied in Session B.

4.2.3. Effectiveness of nonverbal acoustic information

To verify hypothesis 2, we analyzed the relationship between users' talking motivation and nonverbal acoustic information. Simple nonverbal acoustic features were extracted from user speech during dialog and the correlation between the average value of nonverbal acoustic features for four turns before the topic switched and the user's talking motivation on the topic switch was calculated. The correlation values between the nonverbal acoustic features and the user's talking motivation are shown in Table 2. A certain degree of correlation was confirmed for speech length and fundamental frequency, demonstrating hypothesis 2. However, this information is still insufficient to control the timing of switching topics. In the future, we will consider methods, such as combining multiple features to make decisions of switching topics.

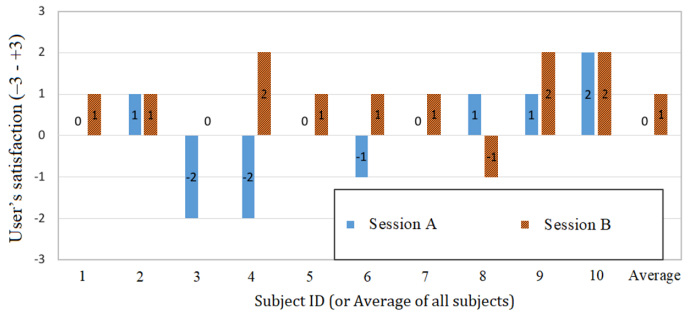


Figure 9. Evaluation results of user's satisfaction

Table 2. Correlation between acoustic features and user's motivation

Features	Correlation
Time from ending of the system's speech to beginning user's speech	0.036
Average volume of the speech interval	-0.044
Speech length	0.29
Tone ratio	-0.14
Fundamental frequency	0.37

5. CONCLUSION

In this paper, we proposed a topic induction method using users' talking motivation to automatically estimate the user's talking motivation from nonverbal acoustic information, to improve the efficiency of gathering information by a chat dialog system. In the automatic estimation of the user's talking motivation, results showed that the user's talking motivation varied depending on the interest level in the current topic, correlating to several nonverbal acoustic information. Additionally, we compared the estimation error among several estimation methods and confirmed the error reduction using the information of multiple turns. In the proposed topic induction method, the user's talking motivation is used as input, and a dialog experiment with a dialog system that transitions from the current topic to either a near or far topic is conducted using the Wizard of Oz method. Thus, the system that transitions to a topic close to the current topic when the user's talking motivation is high, and a far topic otherwise, recorded higher user satisfaction. Furthermore, the user's talking motivation was weakly correlated with the nonverbal acoustic information obtained from the user's speech. In the future, it will be necessary to automatically estimate the user's talking motivation using nonverbal acoustic information from multiple turns and to verify such estimation using an automated system that switches topics toward high user satisfaction.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP16K01543.

REFERENCES

[1] I. Azimi, A. M. Rahmani, P. Liljeberg, and H. Tenhunen, "Internet of things for remote elderly monitoring: a study from user-centered perspective," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, pp. 273–289, 2017. doi: 10.1007/s12652-016-0387-y.

[2] S. Majumder, E. Aghayi, M. Noferesti, H. Memarzadeh-Tehran, T. Mondal, Z. Pang, and M. J. Deen, "Smart homes for elderly healthcarerecent advances and research challenges," *Sensors*, vol. 17, no. 11, p. 2496, 2017. doi: 10.3390/s17112496.

[3] Y. Kobayashi, D. Yamamoto, T. Koga, S. Yokoyama, and M. Doi, "Design targeting voice interface robot capable of active listening," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2010, pp. 161–162. doi: 10.1109/HRI.2010.5453214.

[4] Y. Sakai, Y. Nonaka, K. Yasuda, and Y. I. Nakano, "Listener agent for elderly people with dementia," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (HRI)*, 2012, pp. 199–200. doi: 10.1145/2157689.2157754.

[5] H. Nagasaka, H. Kawanaka, K. Yamamoto, K. Suzuki, H. Takase, and S. Tsuruoka, "A study on topic control method for robot-assisted therapies dementia evaluation using simple conversation with robots ," *Transactions of Japanese Society for Medical and Biological Engineering*, vol. 51 Supplement, pp. R–262, 2013. doi: 10.11239/jsmbe.51.R-262.

[6] O. Yoshito, M. Makoto, and K. Shirai, "Construction of decision model for the spoken dialogue system to close communication," *IPSJ SIG Technical Report*, 2011-HCI-142(2), pp. 1–8, 2011.

[7] T. Ishihara, K. Nitta, F. Nagasawa, and S. Okada, "Estimating interviewee's willingness in multimodal human robot interview interaction," in *Proceedings of the 20th International Conference on Multimodal Interaction (ICMI)*, 2018, pp. 1–6. doi: 10.1145/3281151.3281153.

[8] T. Meguro, R. Higashinaka, Y. Minami, and K. Dohsaka, "Controlling listening-oriented dialogue using partially observable markov decision processes," in *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference*, 2010, vol. 2, pp. 761–769. doi: 10.1145/2513145.

[9] P. Lison, "A hybrid approach to dialogue management based on probabilistic rules," *Computer Speech & Language*, vol. 34, no. 1, pp. 232–255, 2015. doi: 10.1016/j.csl.2015.01.001.

[10] P. Lison and C. Kennington, "OpenDial: A toolkit for developing spoken dialogue systems with probabilistic rules," in *54th Annual Meeting of the Association for Computational Linguistics (ACL) 2016 - System Demonstrations*, 2016, pp. 67–72. doi: 10.18653/v1/P16-4012.

[11] S. Yokoyama, D. Yamamoto, Y. Kobayashi, and M. Doi, "Development of dialogue interface for elderly people-switching the topic presenting mode and the attentive listening mode to keep chatting," in *IPSJ SIG Technical Report*, vol. 2010-SLP-80, no. 4, pp. 1–6, 2010.

[12] B. Schuller, R. Müller, and B. Hörnler, A. Höethker, H. Konosu, G. Rigoll, "Audiovisual recognition of spontaneous interest within conversations," in *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces*, 2007, pp. 30–37. doi: 10.1145/1322192.1322201.

[13] Y. Chiba, T. Nose, and A. Ito, "Analysis of efficient multimodal features for estimating user's willingness to talk: comparison of human-machine and human-human dialog," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017, pp. 428–431. doi: doi: 10.1109/APSIPA.2017.8282069.

[14] N. Saito, S. Okada, K. Nitta, Y. I. Nakano, and Y. Hayashi, "Estimating user's attitude in multimodal conversational system for elderly people with dementia," in *2015 AAAI spring symposium series*, 2015, vol. SS-15-07, pp. 100–103.

[15] M. Henderson, B. Thomson, and J. D. Williams, "The second dialog state tracking challenge," in *SIGDIAL 2014 - 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference*, 2014, pp. 263–272. doi: 10.3115/v1/W14-4337.

[16] M. Henderson, B. Thomson, and S. Young, "Word-based dialog state tracking with recurrent neural network," in *SIGDIAL 2014 - 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference*, 2014, pp. 292–299. doi: 10.3115/v1/W14-4340.

[17] K. Yoshino, T. Hiraoka, G. Neubig, and S. Nakamura, "Dialogue state tracking using long short term memory neural networks," in *Proceedings of Seventh International Workshop on Spoken Dialog Systems*, 2016, pp.1-8.

[18] N. M. Fraser and N. Gilbert, "Simulating speech systems," *Computer Speech & Language*, vol. 5, no. 1, pp. 81–99, 1991. doi: 10.1016/0885-2308(91)90019-M.

[19] M. Okamoto, Y. Yang, and T. Ishida, "Wizard of Oz method for learning dialog agents," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2001, vol. 2180, pp. 20–25. doi: 10.1007/3-540-44799-7_3.

[20] S. Fujie, D. Yagi, Y. Matsusaka, H. Kikuchi, and T. Kobayashi, "Spoken dialogue system using prosody as paralinguistic information," in *Proceedings of Speech Prosody*, 2004, pp. 387–390.

[21] T. Ohsuga, M. Nishida, Y. Horiuchi, and A. Ichikawa, "Investigation of the relationship between turn-taking and prosodic features in spontaneous dialogue," in *Proceedings of Interspeech*, 2005, pp. 33–36. doi: 10.21437/Interspeech.2005-32.

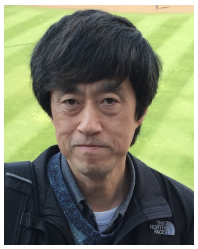
[22] H. Isahara, F. Bond, K. Uchimoto, M. Utiyama, and K. Kanzaki, "Development of Japanese WordNet," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 2008.
[23] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, 2013, pp. 3111–3119.
[24] F. Eyben, M. Wöllmer, and B. Schuller, "opensmile–the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462. doi: 10.1145/1873951.1874246.
[25] B. Schuller, S. Steidl, and A. Batliner, "The interspeech 2009 emotion challenge," in *Proceedings of Interspeech*, 2009, pp. 312–315. doi: 10.21437/Interspeech.2009-103.
[26] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Dept. Comp. Sci., University of Waikato, Hamilton, NewZealand, 1999.

BIOGRAPHIES OF AUTHORS

Ryota Togai was a master-course student at department of informatics, Graduate School of Integrated Science and Technology, Shizuoka University. He obtained bachelor's degree in informatics from Shizuoka University in 2016, and master's degree in Informatics from Shizuoka University in 2018. He can be contacted at email: togairyota@gmail.com.

Takashi Tsunakawa is a lecturer at the faculty of informatics, Shizuoka University from 2019. He obtained a master's degree and Ph.D in information science and technology at the University of Tokyo in 2005 and 2010, respectively. He was a project researcher at the University of Tokyo, a scientific researcher, and an assistant professor at Shizuoka University. His researches are in the fields of natural language processing, especially in machine translation, dialog systems, and assistance in education. He is affiliated with ACL, the association for natural language processing in Japan, the information processing society of Japan, and the Japanese society for artificial intelligence. Besides, he is also involved in some groups including a SIG for patent translation. Further info on his homepage: https://wwp.shizuoka.ac.jp/tsunakawa/. He can be contacted at email: tuna@inf.shizuoka.ac.jp.

Masafumi Nishida is an associate professor at the faculty of informatics since 2015, Shizuoka University. He completed a Ph.D. in engineering at Ryukoku University in 2002. He was an assistant professor at Chiba University, an associate professor at Doshisha University, and a designated associate professor at Nagoya University. His researches are in the fields of speech information processing, behavior signal processing, and well-being information technology. He is affiliated with the information processing society of Japan, human interface society, the institute of electronics, information and communication engineers in Japan, the acoustical society of Japan, and the Japanese society for artificial intelligence. Further info on his homepage: https://lab.inf.shizuoka.ac.jp/nisimura/Nishida.html. He can be contacted at email: nishida@inf.shizuoka.ac.jp.



Masafumi Nishimura is a professor at the faculty of informatics, Shizuoka University since 2014. He obtained a master's degree at the graduate school of engineering science, Osaka University in 1983. He obtained Ph.D. in engineering. He was engaged in research on speech-language information processing at IBM research Tokyo. He received the Yamashita memorial research award from the information society of Japan in 1998, and the technical development award from the acoustical society of Japan in 1999. His researches are in fields of speech information processing, human augmentation using sound information, assistance for the elderly and the disabled. He is affiliated with IEEE, the information processing society of Japan, the institute of electronics, information and communication engineers in Japan, the acoustical society of Japan, and the Japanese society for artificial intelligence. Further info on his homepage:https://lab.inf.shizuoka.ac.jp/nisimura/. He can be contacted at email: nisimura@inf.shizuoka.ac.jp.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 23~33
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp23-33

Karawitans' musician brain adaptation: standardized low-resolution electromagnetic tomography study

Indra K. Wardani[1], Phakkharawat Sittiprapaporn[2], Djohan[3], Fortunata Tyasinestu[1,4]
[1]Department of Music Education, Faculty of Performing Arts, Indonesia Institute of the Arts, Yogyakarta, Indonesia
[2]Neuropsychological Research Laboratory, Department of Anti-Aging and Regenerative Science, School of Anti-Aging and Regenerative Medicine, Mae Fah Luang University, Bangkok, Thailand
[3]Department of Performing Arts, Faculty of Performing Arts, Indonesia Institute of the Arts, Yogyakarta, Indonesia
[4]Graduate School of the Indonesia Institute of the Arts, Yogyakarta, Indonesia

Article Info

Article history:

Received Aug 31, 2021
Revised Jul 9, 2022
Accepted Aug 7, 2022

Keywords:

Brain
Electroencephalography
Karawitan
Music
Standardized low-resolution electromagnetic tomography

ABSTRACT

The rapid advancement of music studies has resulted in a plethora of multidisciplinary participants. Rather than distinguishing between musicians and non-musicians' brain activity, the current study indicated differences in brain activity while musicians listened to music based on their musical experience. In Go/NoGo response task reaction times, it showed that effects between treatments and visits were different across periods of cognitive function tests. The cognitive function at post-listening assessment out-performed the pre-listening in terms of reaction times 531.94 (±24.70) msec for post-listening assessment; and 557.13 (±37.15) msec for pre-listening assessment. The results of using electroencephalography (EEG) recording in an experimental manner with Karawitan musicians (N=20) revealed that listening to unknown cultural music, Mozart's Piano Sonata in C Major, and western music resulted in increased brain activity. Furthermore, while Karawitan musicians were listening to Mozart's Piano Sonata in C Major, the major brain activity occurred in the frontal lobe. This outcome will elicit additional consideration of music's integration, such as neuroscience of music.

Corresponding Author:

Phakkharawat Sittiprapaporn
Neuropsychological Research Laboratory, Department of Anti-Aging and Regenerative Science, School of Anti-Aging and Regenerative Medicine, Mae Fah Luang University
87-88 P.S. Tower 25th Floor, Asoke Montri Road, Khlongtoey Nua, Wattana, Bangkok 10110, Thailand
Email: wichian.sit@mfu.ac.th

1. INTRODUCTION

Music and its cognition have long been investigated by neuroscience, a cognition-focused field. It has attempted to explain not just the process of musical perception, but also the anatomical and functional impact of music on the brain. Lots of research has gone into how musical practice affects cognitive function. Passively listening to music while doing something increases cognitive performance [1]. Some music tempo and mode treatments boost spatial cognition by enhancing arousal and mood [2]. Outside of music, musical training enhances brain function and cognitive abilities [3]. Previous research on symphonic musicians and non-players used the visuospatial task to show how complexity of musical training influences activation of Broca's region during the test, which enhanced performance [3]. Gaser and Schlaug examined professional, amateur, and non-musicians to see how their brains differed [4].

Previous research suggested that long-term musical practice by amateur and professional musicians was the main reason. In a previous study [1], [5], music experience was used as a form of previous

conditioning to find the people who took part [1], [5]. Involvement in long-term musical training and skill acquisition allows for a different psychological process. For example, musicians must memorize musical expressions, improvise music, and recognize a note without a referential note [5]. Each of these challenges in music practice was viewed as a brain stimulation to improve performance and perception. Musical experience was defined as "having received" or "having not received" musical instruction. It is easy to distinguish between the subject and the potential difference that occurred.

As a result, rather than looking at distinctions between musicians and non-musicians, the current study tried to explain the brain activity of Karawitan musicians listening to Mozart's Piano Sonata in C Major. A previous study on musical preference and cognitive style found that people with various cognitive styles have certain personality traits, and musical genre became a distinct variable. We investigated how Karawitan artists with diverse musical backgrounds perceive acoustic cues. An electroencephalogram (EEG) frequency range and standardized low-resolution electromagnetic tomography (sLORETA) were used to quantify cortical activation and locate the brain region contributing to the scalp recorded auditory inputs. The current study's goal was to investigate the lateralization of musical experiences and cognitive function using Mozart's Piano Sonata in C Major.

2. RESEARCH METHOD

2.1. Participants

This study included healthy right-handed people with normal hearing and no known neurological problems. This study involved 20 Karawitan musicians aged 23-29 (mean 28.25±1.41). All Karawitan musicians actively learned Gendhing Lancaran practical music lessons and did not learn piano music lessons for three years. The length of time spent learning music was considered. All participants spoke Bahasa Indonesia as their primary language. They had normal hearing and eyesight, and their health conditions were confirmed via physical examination. The procedure had been described and authorized by everybody. The Graduate School of the Indonesia Institute of the Arts, Yogyakarta, Indonesia, approved the experiment in accordance with the 1964 Helsinki Declaration and its following updates. The flowchart for study enrollment and completion as well as the timeline of the study are shown in Table 1 and Figure 1.

Table 1. The demographic data of participants

Characteristics	Participants (n = 20)
Age (years), mean (SD)	28.25±1.41
Male	6
Female	14
Education (level)	
1st year	8
2nd year	4
3rd year	6
4th year	2

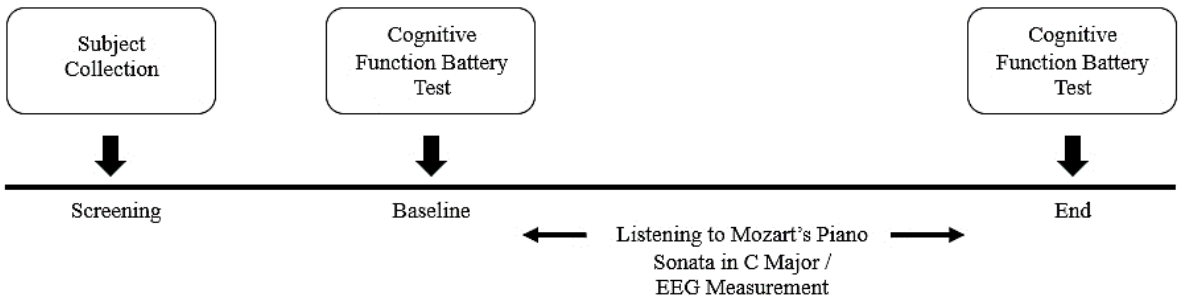


Figure 1. The timeline of both cognitive function battery test and and EEG measurement in the study

2.2. Study design, task and cognitive function assessment

Prior to beginning, the protocol was approved by the Graduate School of the Indonesia Institute of the Arts, Yogyakarta, Indonesia. This study was carried out in accordance with the Helsinki Declaration. Informed consent from all participants was provided before being enrolled in the investigation. Upon enrollment, participants were assigned identification numbers in ascending order and randomly assigned. A

computerized psychological battery was used to evaluate each participant's cognitive abilities at baseline, prior to and after listening to stimuli. The participants' unfamiliar music was Mozart's Piano Sonata in C Major, which served as a stimulus. The sound lasted one minute. It was delivered binaurally through headphones at 85 dB sound pressure level (SPL). The beginning of stimuli was used to time-lock the EEG signal recording. Participants were instructed to focus on the stimuli delivered through earbuds. The cognitive battery of psychometric and psychological tests was administered via a computer interface. This study's cognitive battery included memory tasks, i.e., the Go/NoGo test. The exam was used to assess working memory updating, shifting, and inhibition. The participants were assessed by research assistants who were either graduate students or degree holders in music and/or psychology from the Indonesia Institute of the Arts of Yogyakarta, Indonesia. Each test session's accuracy and response times were recorded and expressed as a percent and milliseconds. Participants were instructed to memorize a set of X and O letters to test their cognitive performance. For this test, a list of X and O letters was presented to the participant one at a time, and the participant was asked to memorize the X and O letters. The participant was required to recall them when they were asked to select only X, but not O. During this task, X and O, which served as Go and No-Go signals, were presented on a monitor at a distance of around 150 cm from the participants' eyes. The X represented the 'Go' condition with 80 percent probability, and the O represented the 'NoGo' condition, with 20 percent probability. The task consisted of 200 stimuli (2 blocks; 20% NoGo signals). In the task, the alphabets X was used as Go and O as NoGo signals. Participants were required to hold their reactions to a single NoGo (O) letter and a series of Go (X) stimuli. At the end of each block, participants were given feedback. The reaction buttons were placed under the participants' palms in a soundproofed and electrically protected chamber. Participants had to press the response pad as quickly as they could (with their dominant hand) every time the more frequent X (Go) stimulus appeared on the computer screen, and to withhold their reactions to the less frequent O (NoGo) stimulus. The order of conditions was counterbalanced across participants as sown in Figure 2.

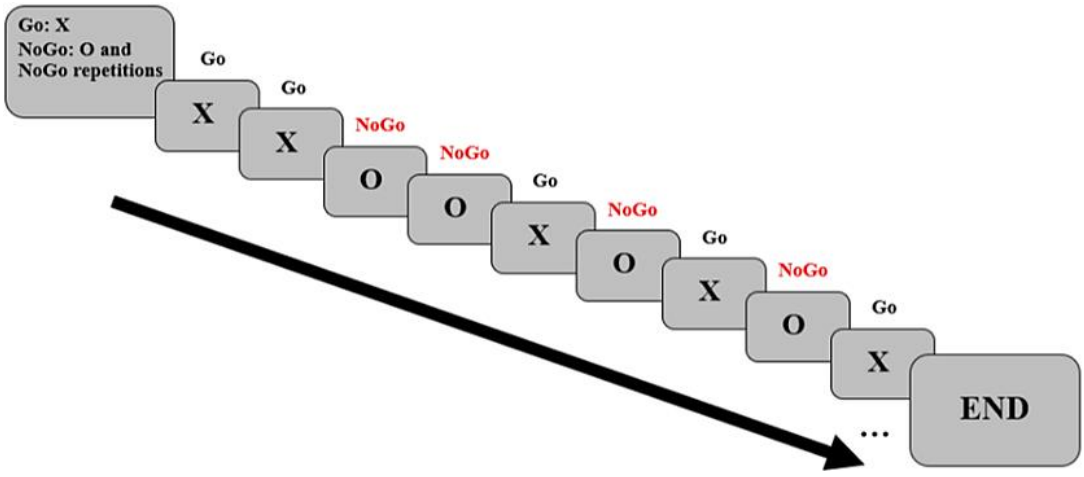


Figure 2. Parametric Go/NoGo task

2.3. Behavioral recording and analyses

The cognitive function challenge required subjects to press buttons accurately and quickly. So, button presses were classified as correct (button code matched stimulus type), incorrect (button code did not match stimulus type), or missed (no button press). The difference in timing between the button codes and the stimulus start was used to calculate reaction times. Participants' reaction accuracy and times were measured from their key presses for each cognitive function task. Reaction times for correct, incorrect, and missed responses were measured as the difference in timing between the stimulus onset and key press reaction. We only included trials with reaction times of 100–1500 ms. This was done to prevent accidental or excessively delayed button pushing due to attention or cognitive exhaustion. Each participant's reaction times were averaged. Reaction times were averaged and corrected for learning affects over time across trials for each participant. On significant analysis of variance (ANOVA) results, we applied Tukey's post hoc analysis. The cognitive function exam used student t-tests for accuracy and reaction speeds. The *p*-value of 0.05 was chosen as the statistical significance level [6].

2.4. Electroencephalographic recording procedure

EEG recordings were based on signals detected through the scalp with a wearable, multi-electrode array neuroheadset (EMOTIV Epoc Plus, San Francisco, USA). The electrical activity of 14 active electrodes (AF3, F3, F7, FC5, T7, P7, O1, O2, P8, T8, FC6, F8, F4, and AF4) as shown in Figure 3(a) - Figure 3(b) was recorded according to the International 10-20 Electrode Positioning System. The left and right mastoids were used as reference electrodes. Manual reference electrodes were placed on ipsilateral mastoids (M1 and M2), with Fp1 and Fp2 electrodes employed for ocular artifact detection. EEGs were 30,000 times amplified and filtered at 0.1-100 Hz. Eye movement and muscular artifacts were extensively analyzed after filtering. Epochs with voltage variations of over 100 µV in any EEG channel were excluded. All responses were recalculated offline against an average reference for additional examination as shown in Figure 3(c). The resistance of the electrodes was less than 10 kΩ. With a 0.05 to 100 Hz band pass, the EEG signals were amplified, captured at 500 Hz, and the live signal data was saved to a hard disk for off-line processing. A 0.1–30 Hz band pass was then used to digitally off-line filter the recorded EEGs. The epoch on which the average was calculated was 500 milliseconds for the commencement of the presenting stimuli. All neural and ocular artifacts were removed from the continuous EEG prior to the extraction of EEG waves. The baseline correction was also applied to each epoch, with any changes in voltage 0.1 µV or 70 µV rejected from further analysis. After registration, the data was re-referenced offline to the common average montage, followed by correction and rejection of artifacts. EEG epochs with absolute amplitudes greater than 100µV were automatically flagged and removed from further investigation. Before averaging, all channels were subjected to artifact rejection with a threshold of ± 100 µV. The total recording time was 5 minutes for each of the cognitive tests. All EEG analyses were performed using TestBench analysis software (EMOTIV Epoc Plus, San Francisco, USA), featuring source reconstruction, signal analysis, and MRI processing tools by sLORETA analysis software.

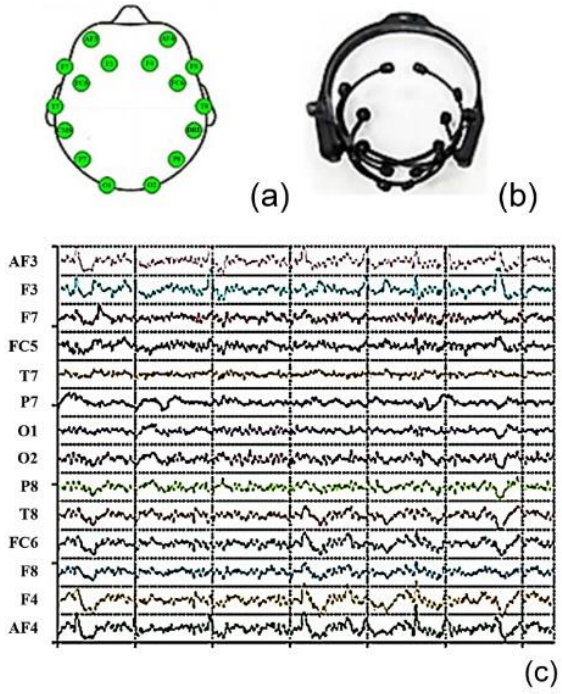


Figure 3. Electroencephalographic recordings were based on signals detected through the scalp with a wearable, multi-electrode array neuroheadset according to the International 10-20 Electrode Positioning System; (a) electroencephalographic device (EMOTIV Epoc Plus), (b) the 14-channels electrode montage, and (c) all electrical activity responses were recalculated offline against an average reference.

2.5. Data pre-processing

The EEG data was converted from.edf to.csv and then captured by the TestBench application. We deleted unnecessary tables and ensured that collected data was consistent with geographical analysis in this Excel-compatible file. The electrical activities of Karawitan musicians' brains while listening to Mozart's

Piano Sonata in C Major were characterized as a moment of global field power (GFP) and were segmentally separated for each of the frequency ranges, for example, delta, theta, alpha, and beta, separately, using stable scalp-potential topography [6]. An EEG map's spatial standard deviation of all voltage values equaled the GFP peak. Each participant's mean GFP peak amplitude was calculated, and each participant's mean spontaneous EEG map was checked for artifacts. These were then averaged across all subjects. The brain's current source density distribution was evaluated using sLORETA, which was added to the electrical scalp field [7]. The smoothest of all possible source configurations over the brain volume was identified using the entire squared Laplacian [7]. A brain electric field map's electric strength (hilliness) was assessed using the GFP peak measure an EEG map's spatial standard deviation of all voltage estimates. The GFP peak measure is higher on a mountainous map than on a flat map. This GFP was self-contained [8], [9]. As a result, the spatial standard deviation of global field power provides a reference-independent descriptor of the potential field. The latencies of evoked potential components are determined by global field power maxima. Global field power builds up with time [10]. Students learned about global field power computation, component latency, global dissimilarity of potential field distributions, and topographical temporal segmentation using multichannel data. We got GFP by averaging the EPs across all scalp channels except electrooculographic. Each person's mean GFP peak amplitudes were obtained. Their GFP peak amplitudes were also computed [7]–[10]. The GFP waveforms were examined using cognitive function tests. Brain responses to the tango piece measured with electroencephalography (EEG).

2.6. Statistical and data analysis

While listening to Mozart's Piano Sonata in C Major, Karawitan musicians' brainwaves were continuously monitored. The smoothest source configurations across the brain volume were produced by limiting the absolute squared Laplacian of source quality. Quantitative data is provided as means with standard deviations. The data were analyzed using SPSS Program (Mae Fah Luang University) version 21.0, renewal quote number: 26500879; Passport advantage site number: 3547818. To determine the effects of music on cognitive function over the time periods (before and after listening), mean response times (RTs) and correct responses, as well as cognitive function analyses, were performed using a two-way paired t-test. Response times to cognitive function battery tests were measured for correct responses. One-way ANOVAs were performed on accuracy and reaction times for the cognitive function tests. Statistical results were considered significant at $p < 0.05$ [6].

3. RESULTS AND DISCUSSION

3.1. Cognitive enhancing effects

In Go/NoGo response task reaction times, the two-way (treatment x assessment) interaction effect was statistically significant, with $F(1,19) = 19.37$, $p < 0.001$, indicating that the two-way interaction effects between treatments and visits were different across periods of cognitive function tests. Baseline and end-of-treatment reaction times scores are reported in Table 2. The cognitive function at post-listening assessment out-performed the pre-listening in terms of reaction times (531.94 (±24.70) msec for post-listening assessment; and 557.13 (±37.15) msec for pre-listening assessment.

Table 2. The effect of Piano Sonata in C Major on the cognitive function battery test

Cognitive Function Test[α]	Score at Pre-listening	Score at Post-listening	*p*-value
Go/NoGo Test			
(%) Accuracy (Go)	98.75	99.31	
(%) Error (Go)	1.25	0.69	
Response time (ms) (Go) (mean±SD)	557.13 (±37.15)	531.94 (±24.70)	< 0.05*
(%) Accuracy (NoGo)	87.64	89.37	
(%) Error (NoGo	12.36	10.63	

[α] Test parameters: Go/NoGo Accuracy: Go/NoGo response task accuracy scores; Go/NoGo Error: Go/NoGo response task incorrect and omission scores; Go/NoGo response time: Go/NoGo response task mean reaction time in milliseconds (ms). * *p* value < 0.05.

3.2. Electroencephalographic data

Table 3 shows the effect of Mozart's Piano Sonata in C Major listening assessed by the electroencephalography. GFP was plotted as a function of time, and the occurrence times of GFP maxima were used to determine each frequency band sensitivity. The grand mean GFP peak amplitude of each frequency band over subjects is shown according to the experimental setting. The electrical activities of

Karawitan musicians' brains computed by sLORETA showed that alpha wave had the highest electrical activity (9.705±0.12µV, $t(19) = 3.06$; $p < 0.05$) compared to other waves (e.g., delta wave: 2.529±0.25µV, $t(19) = 2.18$; $p < 0.05$, theta wave: 2.58±0.61µV, $t(19) = 3.66$; $p < 0.05$, and beta wave: 2.621±0.45µV, $t(19) = 2.19$; $p < 0.05$), respectively, while tuning in to Mozart's Piano Sonata in C Major, as shown in Table 3 and Figure 4.

Table 3. Electrical activities (µV) of Karawitan Musicians' brains while tuning in to Mozart's Piano Sonata in C Major as computerized by sLORETA

Frequency Bands	Mozart's Piano Sonata in C Major Mean (±SD)
delta	2.529 (±0.25)
theta	2.580 (±0.61)
alpha	9.705 (±0.12)
beta	2.621 (±0.45)

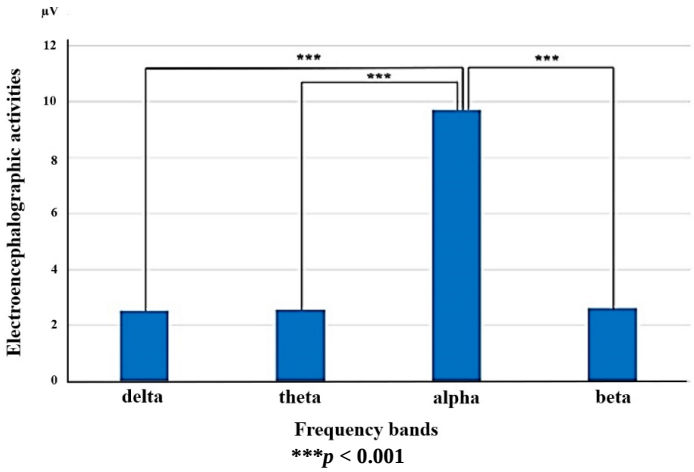


Figure 4. Electrical activities (µV) (Mean±SD) of Karawitan musicians' brains while tuning in to Mozart's Piano Sonata in C Major as computerized by sLORETA

3.3. Source localization data

Source localization analyses were performed utilizing sLORETA [8]. Figure 2 shows the *xyz*-values in Talairach space as determined with the sLORETA. The graphical representation of sLORETA *t*-statistic while Karawitan musicians listening to Mozart's Piano Sonata in C Major at superior frontal gyrus (STG)-frontal lobe (Brodmann area 10; X = -10, Y = 60, Z = 30; MNI coords; Best match at 0 mm; 9.71 µV) in the right-hemisphere (RH). The yellow color indicates local maxima of increased electrical activity in the right hemisphere (middle-to-back region) through the reference brain. A blue dot marks the center of significantly increased electric activity of an alpha wave as shown in Figure 5.

When participants listened to Mozart's Piano Sonata in C Major, which was not related to their own culture, western music, the frontal lobe was the place where dominant brainwave (i.e. alpha wave) occurred. Karawitan musicians listened to Mozart's Piano Sonata in C Major for this study. When listening to Mozart's Piano Sonata in C Major, Karawitan musicians had the highest Alpha (α) brainwave activity. This could indicate that when the individuals were unfamiliar with the cultural music they were listening to, their brain activity increased. Individuals' brain activity is increased by familiarity rather than liking, according to a prior study [11]. The previous study indicated that various emotion-related areas such as the amygdala, putamen, anterior cingulate cortex, and thalamus were activated during familiar music listening using functional magnetic resonance imaging (fMRI). In our study, participants' brains were stimulated especially in the superior frontal gyrus by unrelated cultural yet familiar music (STG). In a meta-analysis of brain areas activated by familiar music, the left superior frontal gyrus was the most stimulated, followed by the ventral lateral [12]. The frontal gyrus is assumed to be stimulated by semantic memory of familiar music, while the ventral lateral, related to the motor cortex, is supposed to be stimulated by motoric anticipation of familiar music's rhythms [9]. Unable to locate it in our study because movement was not allowed while listening to music, a recent study found the ventral lateral activated by familiar music [12].

Our understanding about music to relax the brain might still be controversy. Take an example from Patston and Tippett's study where expert musicians even struggle harder to do the linguistic task when listening to familiar piano excerpts because of its overlapping processing showed that relaxing and cooling down effects also depend on the musical experience of individuals [1]. However, previous study demonstrated that the more participants familiar into the music, the higher his brain activity which means it is more relaxed. According to our findings, the claim of music relaxation shouldn't be made by putting music randomly as a background without examining an individual's musical background. The brain activity of our participants while listening to unfamiliar music, the Karawitan musicians showed higher brain activity (i.e. alpha wave) while listening to Mozart's Piano Sonata in C Major. Despite the fact that the Piano Sonata was unknown music to Karawitan musicians, their brain activity was greater in the alpha wave while listening to these portions. This circumstance could be explained by two assumptions. The initial assumption goes through the typical process of this type of music. This excerpt was musically distinct from our participants' cultural backgrounds, particularly in terms of speed and melody succession. Mozart's Piano Sonata in C Major, written in allegro, featured quick music with rapid melodic succession. It was considerably different from the traditional music of our participants. For example, Gendhing Lancaran's tempo was relatively slow and quiet, and the melody was a cyclical motif that was not as fast as Mozart's Piano Sonata in C Major. The second hypothesis could be linked to musical perception and experience. Furthermore, a prior study found that musical experience and perception can be influenced by cultural differences [13]. The significance of early musical experience in promoting auditory sequence memory in musicians has also been clarified [14]. The memory tasks of auditory, visual, and audio-visual stimuli were used in this previous study on a wide range of participants, including musicians, gymnasts, video game players, and psychology students. The results revealed that while there was no significant difference in the visual or audio-visual tasks, there was a significant difference in the audio task, where musicians scored higher [14]. Music training and performance have been demonstrated to boost cognitive function in older people. A previous study examined the effects of music on the brain structure of older people. Music training was found to be favorably and significantly connected to the volume of the inferior frontal cortex and parahippocampus. Music training increased volume in the posterior cingulate, insula, and medial orbitofrontal cortex. The study found a relationship between musical actions and executive function, memory, language, and emotion. Because gray matter diminishes with age, this earlier research suggests that musical training may help older people overcome age-related brain volume declines [15].

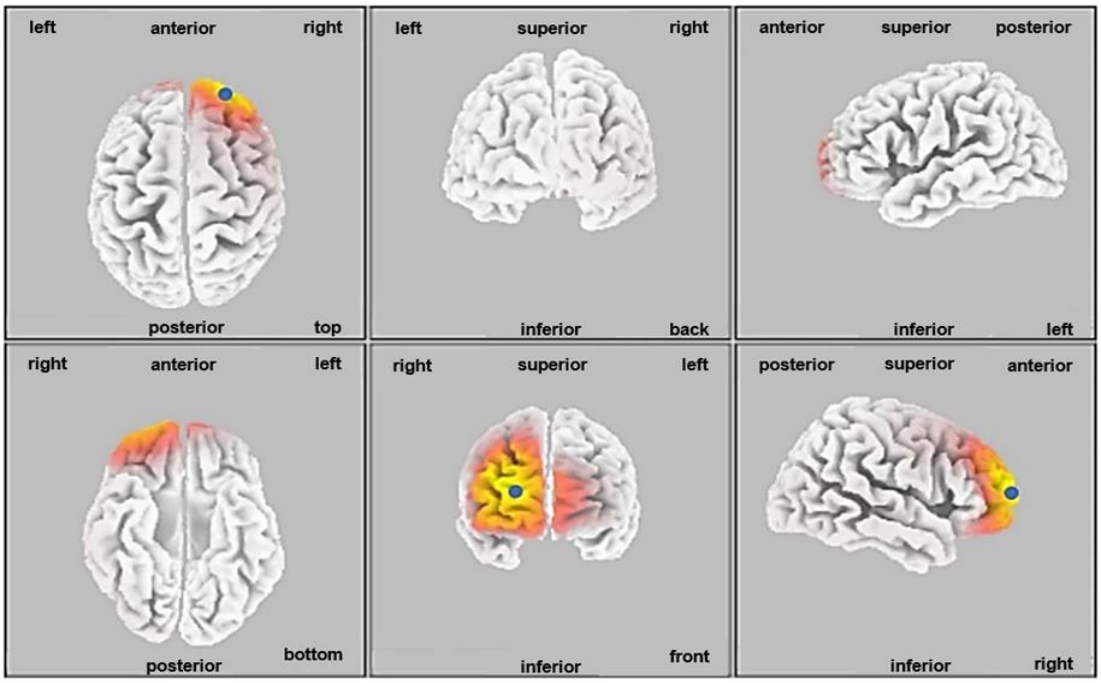


Figure 5. Graphical representation of the sLORETA while Karawitan musicians tune in to Mozart's Piano Sonata in C Major. The yellow color indicates local maxima of increased electrical activity. A blue dot marks the center of significantly increased electric activity of an alpha wave

Listening to music is first and foremost a human experience that becomes aesthetic when the listener totally immerses himself or herself in it. In a recent neuroimaging investigation, a relationship was established between the auditory cortex, the reward brain system, and mind wandering [16]. Music and language use harmonic complex perception. Numerous studies have connected musical training to greater harmonic complex processing. However, the benefit may not be universal across pitch models. Musicality can be reliably linked to objective measures of perception, according to a previous study. Musicianship also influences monotic/diotic and dichotic integration pitch assessments. Collectively, the findings update artists' neurobehavioral profiles and enhance creative capacity assessments [17]. In certain clinical studies, neutral and happy music reduce anxiety. An earlier study found that different emotional music types may have different mechanisms in state anxiety therapy. Neutral music reduces state anxiety. Neutral music was associated with decreased occipital lobe power spectral density and increased occipital-frontal functional connectivity. Happy music reduced state anxiety and enhanced occipital-right temporal functional connections. This earlier investigation may enable future nonpharmaceutical clinical therapy to better grasp state anxiety through music therapy [18].

The most prevalent complication of an acquired brain injury (ABI) is cognitive impairment, which can have a significant influence on a person's life and rehabilitation prospects. When compared to non-musicians, musicians have better cognitive control, attention, and executive functioning as a result of their music training. Music therapy is a technique that employs learning to play an instrument, specifically the piano, to stimulate and rebuild cognitive networks after a brain injury [19]. Music interventions are also viable remedies for Alzheimer's disease symptoms (AD). Music interventions can be active or receptive depending on the subjects' participation. It is possible that separate brain areas are involved in active and receptive music tasks. The clinical benefits of two types of music therapy and a control activity were compared in a recent study. Active music intervention can help with Alzheimer's symptoms and should be given as a supplement to standard care. The data demonstrate that combining AMI with standard treatment can help mild-to-moderate AD patients improve their cognition, behavior, and reliance [20].

Several studies have recently used another brain imaging technique called "Microstate Segmentation" to investigate human brain mechanisms during music listening. Furthermore, EEG is an excellent tool for investigating global states, as well as the briefer states (microstates) imbedded therein, due to its high time resolution [21]. Dimensional complexity proportions can indicate global states. These indicators count dynamic different brain processes to represent gross states like sleep stages [22], [23]. The global states of healthy, depressed, and schizophrenic people differ [24]. Previous studies used EEG frequency spectra and source localization. Less alpha activity and more delta and theta activity were found. The scalp maps of the EEG depict electric potential distributions. Skin maps with long-term quasi-stable potential patterns [21], [25] are microstates. A 60–120 millisecond microstate length is reported by EEG [21], [25]. Distinct spatial distributions of neuronal activity in the brain result in different scalp potentials. As a result, different microstates process data in different ways. "Potential atoms of thinking and emotion" [26], [27]. Microstates come in several forms [28], [29]. The four microstates usually detected in spontaneous resting-state EEG [30] are A, B, C, and D. The number of occurrences, time span, and mean term can be utilized to describe the microstates of the different classes. These characteristics vary greatly between states and tasks. The concept of microstates described in [31] is closely tied to the concept of symbolization in brain recordings. It was discovered by Lehmann that the scalp potential maps have quasi-stable activity lasting tens to hundreds of milliseconds [21], [31], with quick transition between them in event-related potentials (ERPs) and spontaneous EEG time organization. Lehmann provided symbols to the time periods. As indicated in [29], EEG microstate class representation is a valuable data reduction strategy. It's reasonable to believe that each of these microstates is involved in various activities or cognitive tasks [21], [32]. Scientists frequently assess cognitive activities using the microstates framework by mapping scalp potentials and analyzing spatial aspects. In time-domain, the microstates framework successfully measures the human brain electric field [30]. According to a prior study on musical preference and cognitive style, people with specific cognitive styles have a tendency to have certain personality features, and musical genres have become a unique variable.

Almudena Bartolomé-Tomás and colleagues investigated the relationship between traditional musical genre exposure and the memories of Spanish seniors from Murcia. The idea was to see if memories created by listening to rhythms heard as youngsters changed brain activity. The activation of brain areas was discovered using EEG signals. Using spectral power, the researchers discovered significant differences between "memory-evoked" and "non-memory-evoked" classes in the prefrontal cortex's alpha, beta, theta, and gamma frequency bands. The findings shed light on the listener's emotional state during the experiment [33], [34]. The brain's activity in response to memories acquired through music has also been shown. Other studies [33] have discovered comparable memory-forming zones. The experiment used a low-cost brain–computer interface, the Emotiv EPOC+ headset's 14 channels. Using 32 or 64 channels, several

investigations have found significant differences between the prefrontal and frontal-temporal brain areas. Previous research [35] demonstrated that memory recall alters the alpha and theta bands. No one agrees that major changes occur in the beta and gamma bands. Experiments with different groups of subjects revealed variations that cannot be generalized at this time [36], [37]. The shown ability to distinguish between distinct regions helps us to discover and recognize the zones that are activated during the production of recovered memories, as indicated by others [38].

Tseng set out to identify prefrontal cortex brain activity connected with musical choice. Tseng's research focuses on interpreting EEG bands connected with musical liking. Popular songs induced more frontal theta than music with low and moderate preference ratings. The frontal theta is linked to both emotional and cognitive processes, according to frontal theta-cognitive connections. A study published in Psychological Science found that theta and lower alpha in the frontal lobe are effective indicators of both cognitive and mood [39]. Researchers can no longer study how musicians' emotions are processed in the brain. Their research involved having musicians perform a simple piano piece while adjusting their manner of play to transmit opposing feelings, and self-rating the emotion portrayed on arousal and valence scales. In both distressed and comfortable playing, EEG activity differed [40]. Electroencephalograms (EEGs) are widely used to record brain responses to brief, repeating stimuli. In real life, acoustic impulses are continually blended and cannot be isolated as in music. Because music's acoustic qualities are constantly fluctuating in this aural context, substantial values of various features might occur almost simultaneously. The results of a statistical analysis of the N100 and P200 times and delays corroborate this idea. The responses are more pronounced when these features appear combined, such as brightness and root mean square (RMS) or brightness and spectral flux. Together, RMS and spectral flux give greater reactions than when used alone [41]. Last but not least, music information retrieval algorithms can detect time points in music recordings that correspond to brain reactions. But it's unknown how the music's structure and aural qualities affect the brain's reaction. Haumann and colleagues tested a new method for automatically identifying brain reaction times. They used an existing library of EEG and Magnetoencephalography (MEG) recordings from 48 healthy listeners. Preliminary findings demonstrate that studying music novelty can help understand brain reactions to realistic music [42].

4. CONCLUSION

When Karawitan musicians listened to Mozart's Piano Sonata in C Major, western music, their brains displayed faster frequency bands, i.e. alpha wave activity. The main brain activity occurred in the frontal lobe of right hemisphere. Rather than distinguishing between musicians and non-musicians' brain activity, the current study indicated differences in brain activity while musicians listened to music based on their musical experience. This finding will lead to more research into the integration of music, such as music neuroscience.

ACKNOWLEDGEMENTS

This study was partially supported by Mae Fah Luang University (Electroencephalogram Laboratory 2019), Thailand. In contributing this study, Indra K. Wardani designed the study; collected data. Djohan supervised the research. Fortunata Tyasinestu supervised the research. Phakkharawat Sittiprapaporn designed the study; conducted the research; statistically analyzed and interpreted the data; obtained funding; draft and critical revision of the manuscript. Phakkharawat Sittiprapaporn is corresponding author. We thank all subjects who participated in this study.

REFERENCES

[1] L. L. M. Patston and L. J. Tippett, "The effect of background music on cognitive performance in musicians and nonmusicians," *Music Perception*, vol. 29, no. 2, pp. 173–183, 2011, doi: 10.1525/mp.2011.29.2.173.

[2] G. Husain, W. F. Thompson, and E. G. Schellenberg, "Effects of musical tempo and mode on arousal, mood, and spatial abilities," *Music Perception*, vol. 20, no. 2, pp. 151–171, 2002, doi: 10.1525/mp.2002.20.2.151.

[3] V. Sluming, J. Brooks, M. Howard, J. J. Downes, and N. Roberts, "Broca's area supports enhanced visuospatial cognition in orchestral musicians," *Journal of Neuroscience*, vol. 27, no. 14, pp. 3799–3806, 2007, doi: 10.1523/JNEUROSCI.0147-07.2007.

[4] C. Gaser and G. Schlaug, "Brain structures differ between musicians and non-musicians," *Journal of Neuroscience*, vol. 23, no. 27, pp. 9240–9245, 2003, doi: 10.1523/jneurosci.23-27-09240.2003.

[5] G. Schlaug, "The Brain of Musicians," *The Cognitive Neuroscience of Music*, 2012, doi: 10.1093/acprof:oso/9780198525202.003.0024.

[6] S. A. Bann and A. T. Herdman, "Event related potentials reveal early phonological and orthographic processing of single letters in letter-detection and letter-rhyme paradigms," *Frontiers in Human Neuroscience*, vol. 10, no. APR2016, pp. 1–13, 2016, doi: 10.3389/fnhum.2016.00176.

[7] R. D. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details," *Methods and Findings in Experimental and Clinical Pharmacology*, vol. 24, no. SUPPL. D, pp. 5–12, 2002.

[8] D. Lehmann, "Principles of spatial analysis," *Methods of Analysis of Brain Electrical and Magnetic Signals: Handbook of Electroencephalography and Clinical Neurophysiology*, vol. 1, pp. 309–354, 1987.
[9] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Segmentation of brain electrical activity into microstates; model estimation and validation," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 7, pp. 658–665, 1995, doi: 10.1109/10.391164.
[10] W. Skrandies, "Global field power and topographic similarity," *Brain Topography*, vol. 3, no. 1, pp. 137–141, 1990, doi: 10.1007/BF01128870.
[11] C. S. Pereira, J. Teixeira, P. Figueiredo, J. Xavier, S. L. Castro, and E. Brattico, "Music and emotions in the brain: familiarity matters," *PLoS ONE*, vol. 6, no. 11, 2011, doi: 10.1371/journal.pone.0027241.
[12] C. Freitas, E. Manzato, A. Burini, M. J. Taylor, J. P. Lerch, and E. Anagnostou, "Neural correlates of familiarity in music listening: a systematic review and a neuroimaging meta-analysis," *Frontiers in Neuroscience*, vol. 12, no. OCT, 2018, doi: 10.3389/fnins.2018.00686.
[13] I. Cross, "Music, cognition, culture, and evolution," in *The Cognitive Neuroscience of Music*, Oxford University Press, 2003, pp. 42–56.
[14] A. T. Tierney, T. R. Bergeson, and D. B. Pisoni, "Effects of early musical experience on auditory sequence memory," *Empirical Musicology Review*, vol. 3, no. 4, pp. 178–186, 2008, doi: 10.18061/1811/35989.
[15] L. Chaddock-Heyman, P. Loui, T. B. Weng, R. Weisshappel, E. McAuley, and A. F. Kramer, "Musical training and brain volume in older adults," *Brain Sciences*, vol. 11, no. 1, pp. 1–16, 2021, doi: 10.3390/brainsci11010050.
[16] M. Reybrouck, P. Vuust, and E. Brattico, "Brain connectivity networks and the aesthetic experience of music," *Brain Sciences*, vol. 8, no. 6, 2018, doi: 10.3390/brainsci8060107.
[17] D. Inabinet, J. De La Cruz, J. Cha, K. Ng, and G. Musacchia, "Diotic and dichotic mechanisms of discrimination threshold in musicians and non-musicians," *Brain Sciences*, vol. 11, no. 12, 2021, doi: 10.3390/brainsci11121592.
[18] B. Huang *et al.*, "The benefits of music listening for induced state anxiety: behavioral and physiological evidence," *Brain Sciences*, vol. 11, no. 10, 2021, doi: 10.3390/brainsci11101332.
[19] C. Jones, "The use of therapeutic music training to remediate cognitive impairment following an acquired brain injury: the theoretical basis and a case study," *Healthcare (Switzerland)*, vol. 8, no. 3, 2020, doi: 10.3390/healthcare8030327.
[20] M. Gómez-Gallego, J. C. Gómez-Gallego, M. Gallego-Mellado, and J. García-García, "Comparative efficacy of active group music intervention versus group music listening in alzheimer's disease," *International Journal of Environmental Research and Public Health*, vol. 18, no. 15, 2021, doi: 10.3390/ijerph18158067.
[21] D. Lehmann, H. Ozaki, and I. Pal, "EEG alpha map series: brain micro-states by space-oriented adaptive segmentation," *Electroencephalography and Clinical Neurophysiology*, vol. 67, no. 3, pp. 271–288, 1987, doi: 10.1016/0013-4694(87)90025-3.
[22] J. Wackermann, D. Lehmann, I. Dvorak, and C. M. Michel, "Global dimensional complexity of multi-channel EEG indicates change of human brain functional state after a single dose of a nootropic drug," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 3, pp. 193–198, 1993, doi: 10.1016/0013-4694(93)90007-I.
[23] W. Szelenberger, J. Wackermann, M. Skalski, S. Niemcewicz, and J. Drojewski, "Analysis of complexity of EEG during sleep," *Acta Neurobiologiae Experimentalis*, vol. 56, no. 1, pp. 165–169, 1996.
[24] C. J. Stam, E. M. Hessels-Van Der Leij, J. Meulstee, and J. H. R. Vliegen, "Changes in functional coupling between neural networks in the brain during maturation revealed by omega complexity," *Clinical EEG and Neuroscience*, vol. 31, no. 2, pp. 104–108, 2000, doi: 10.1177/155005940003100209.
[25] D. Lehmann, W. K. Strik, B. Henggeler, T. Koenig, and M. Koukkou, "Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. visual imagery and abstract thoughts," *International Journal of Psychophysiology*, vol. 29, no. 1, pp. 1–11, 1998, doi: 10.1016/S0167-8760(97)00098-6.
[26] D. Lehmann, "Brain electric microstates and cognition: the atoms of thought," *Machinery of the Mind*, pp. 209–224, 1990, doi: 10.1007/978-1-4757-1083-0_10.
[27] D. Lehmann, "Consciousness: microstates of the brain's electric field as atoms of thought and emotion," *The Unity of Mind, Brain and World*, pp. 191–218, 2014, doi: 10.1017/cbo9781139207065.007.
[28] T. Koenig *et al.*, "Millisecond by millisecond, year by year: normative EEG microstates and developmental stages," *NeuroImage*, vol. 16, no. 1, pp. 41–48, 2002, doi: 10.1006/nimg.2002.1070.
[29] C. M. Michel and T. Koenig, "EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review," *NeuroImage*, vol. 180, pp. 577–593, 2018, doi: 10.1016/j.neuroimage.2017.11.062.
[30] T. Koenig, D. Lehmann, M. C. G. Merlo, K. Kochi, D. Hell, and M. Koukkou, "A deviant EEG brain microstate in acute, neuroleptic-naive schizophrenics at rest," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 249, no. 4, pp. 205–211, 1999, doi: 10.1007/s004060050088.
[31] A. Tzovara, M. M. Murray, C. M. Michel, and M. De Lucia, "A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials," *Developmental Neuropsychology*, vol. 37, no. 6, pp. 518–544, 2012, doi: 10.1080/87565641.2011.636851.
[32] D. Lehmann, R. D. Pascual-Marqui, W. K. Strik, and T. Koenig, "Core networks for visual-concrete and abstract thought content: A brain electric microstate analysis," *NeuroImage*, vol. 49, no. 1, pp. 1073–1079, 2010, doi: 10.1016/j.neuroimage.2009.07.054.
[33] A. Fernández-Soto, A. Martínez-Rodrigo, J. Moncho-Bogani, J. M. Latorre, and A. Fernández-Caballero, "Neural correlates of phrase quadrature perception in harmonic rhythm: an EEG study using a brain-computer interface," *International Journal of Neural Systems*, vol. 28, no. 5, 2018, doi: 10.1142/S012906571750054X.
[34] A. Fernández-Sotos, A. Fernández-Caballero, and J. M. Latorre, "Influence of tempo and rhythmic unit in musical emotion regulation," *Frontiers in Computational Neuroscience*, vol. 10, no. AUG, 2016, doi: 10.3389/fncom.2016.00080.
[35] L. Ros *et al.*, "Differences in brain activation between the retrieval of specific and categoric autobiographical memories: an EEG study," *Psicologica*, vol. 38, no. 2, pp. 347–363, 2017.
[36] S. Koelstra *et al.*, "DEAP: a database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012, doi: 10.1109/T-AFFC.2011.15.
[37] K. Subramaniam *et al.*, "Beta-band activity in medial prefrontal cortex predicts source memory encoding and retrieval accuracy," *Scientific Reports*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-43291-7.
[38] A. Bartolomé-Tomás, R. Sánchez-Reolid, B. García-Martinez, A. Fernández-Sotos, and A. Fernández-Caballero, "Memory retrieval in ageing adults through traditional music genres—an experiment based on electroencephalography signals," in *13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019*, Nov. 2019, p. 33, doi: 10.3390/proceedings2019031033.

[39] K. C. Tseng, "Electrophysiological correlation underlying the effects of music preference on the prefrontal cortex using a brain–computer interface," *Sensors*, vol. 21, no. 6, pp. 1–13, 2021, doi: 10.3390/s21062161.
[40] J. E. Pousson *et al*., "Spectral characteristics of eeg during active emotional musical performance," *Sensors*, vol. 21, no. 22, 2021, doi: 10.3390/s21227466.
[41] L. J. Tardón, I. Rodríguez-Rodríguez, N. T. Haumann, E. Brattico, and I. Barbancho, "Music with concurrent saliences of musical features elicits stronger brain responses," *Applied Sciences (Switzerland)*, vol. 11, no. 19, 2021, doi: 10.3390/app11199158.
[42] N. T. Haumann, M. Kliuchko, P. Vuust, and E. Brattico, "Applying acoustical and musicological analysis to detect brain responses to realistic music: a case study," *Applied Sciences (Switzerland)*, vol. 8, no. 5, 2018, doi: 10.3390/app8050716.

BIOGRAPHIES OF AUTHORS

Indra K. Wardani received Bachelor and Master's degree of Music from Music Department, Faculty of Performing Arts, Graduate School of Indonesia Institute of the Arts Yogyakarta, Yogyakarta, Indonesia. She is currently a researcher at Music Department, Faculty of Performing Arts, Indonesian Institute of the Arts, Yogyakarta, Indonesia. Her research interest is Neuroscience of Music. She can be contacted at email: indrakwardani@gmail.com.

Phakkharawat Sittiprapaporn received Bachelor of Arts (Second Class Hons.) in English from Srinakharinwirot University, Thailand, Master of Arts in Linguistics, Institute of Language and Cultural for Research and Development from Mahidol University, Thailand, and Ph.D. in Neurosciences, Neuro-Behavioural Biology Center, Institute of Science and Technology for Research and Development Mahidol University, Thailand. He is currently a Head of Brain Science and Engineering Innovation Research Group, Mae Fah Luang University, and Neuropsychological Research Laboratory, as well as a lecturer at Department of Anti-Aging and Regenerative Science School of Anti-Aging and Regenerative Medicine, Mae Fah Luang University, Bangkok, Thailand. His research interests are cognitive psychology, cognitive neurosciences, cerebral mechanisms in perception and cognition, brain mechanism of music and language perception and cognition, and neurobiology of learning and memory. He can be contacted at email: wichian.sit@mfu.ac.th.

Djohan received Bachelor's degree in Music from Art Institute Music School, Gadjah Mada University, Masters' degree in Psychology from Gadjah Mada University, and Doctoral degree in Psychology from Gadjah Mada University. He is currently a Professor at Music Department, Faculty of Performing Arts, Indonesian Institute of the Arts, Yogyakarta, Indonesia. Recently, He working at Music Performance Department. His research interest is Neuroscience of Music and Arts. He can be contacted at email: djohan.djohan@yahoo.com.

Fortunata Tysrinestu received Bachelor's degree in Music from the Indonesian Institute of the Arts, Yogyakarta and Indonesian Literature from Universitas Gadjah Mada Yogyakarta, Masters' degree in Psychology from Gadjah Mada University, and Doctoral degree in Social Sciences from Gadjah Mada University. She is Associate Professor at Music Department, Faculty of Performing Arts, Indonesian Institute of the Arts, Yogyakarta, Indonesia. She is currently a Director at Graduate School of Indonesia Institute of the Arts Yogyakarta, Yogyakarta, Indonesia. Her research interest includes Music Education and Social Sciences. She can be contacted at email: tyasrin2@yahoo.com.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 34~40
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp34-40

Impedance characteristic of the human arm during passive movements

Md. Mozasser Rahman[1], Ryojun Ikeura[2]
[1]Department of Mechanical Engineering Technology, Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia, Muar, Johor, Malaysia
[2]Department of Mechanical Engineering, Faculty of Engineering, Mie University, Kamihama-1515, Tsu, Mie, Japan

Article Info

Article history:

Received Sep 1, 2021
Revised Jul 1, 2022
Accepted Jul 30, 2022

Keywords:

Human arm
Human-robot cooperation
Impedance characteristics
Minimum jerk trajectory
Passive movements
Single degree-of-freedom

ABSTRACT

This paper describes the impedance characteristics of the human arm during passive movement. The arm was moved in the desired trajectory. The motion was actuated by a 1-degree-of-freedom robot system. Trajectories used in the experiment were minimum jerk (the rate of change of acceleration) trajectories, which were found during a human and human cooperative task and optimum for muscle movement. As the muscle is mechanically analogous to a spring-damper system, a second-order equation was considered as the model for arm dynamics. In the model, inertia, stiffness, and damping factor were considered. The impedance parameters were estimated from the position and torque data obtained from the experiment and based on the "Estimation of Parametric Model". It was found that the inertia is almost constant over the operational time. The damping factor and stiffness were high at the starting position and became near zero after 0.4 seconds.

Corresponding Author:

Md. Mozasser Rahman
Department of Mechanical Engineering Technology, Faculty of Engineering Technology
Universiti Tun Hussein Onn Malaysia
Km 1, Panchor Street, 84600 Pagoh, Muar, Johor Darul Ta'zim, Malaysia
Email: mozasser@uthm.edu.my

1. INTRODUCTION

The human being is the best creature in this universe. In comparison to other creatures, the size and shape of the human body are most favorable in all respect and every movement of the human body is smooth and perfect. In the past, scientists have tried to build a mechanism that imitates parts of the human body to perform the task for mankind. The development of robots during the latter half of the twentieth century is its burning example.

Among the moving parts of the human body, the upper limbs are used most frequently. The main function of the upper limb is grasping and manipulating. This is also used as a walking aid to support the body during gait. The upper limb consists of three main parts, the upper arm, forearm, and hand. It is composed of three chain mechanisms, the shoulder girdle, the elbow, and the wrist, whose association allows a wide range of combined motion. Due to the complexity of the hand mechanism, the wrist was not studied, and the hand was taken as another rigid segment in the extension of the forearm. The movements of the human arm can be divided into three major types: i) active movement, an external force is exerted by the hand; ii) reaching movement, without exerting any external force; and iii) passive movement, a hand is moved by external force [1].

The arm is a multi-joint redundant manipulator. It is found that the normalized speed and velocity profiles for single and multiple joint trajectories are identical [2]. Shoulder velocity profiles remain

unchanged, but the acceleration phase of elbow trajectories is adjusted so that peak velocity and movement time match that of the shoulder joint. They argue that hand-path is the primary movement criterion, and that elbow movement is subordinate to elbow movement in a hierarchical scheme to reduce the available degrees of freedom.

Cruse and Brüwer studied planar reaching movements while recording shoulder, elbow, and wrist angles to determine how subjects solved the redundant degrees of freedom problem [3]. They propose that each limb has an almost comfortable position and that by associating a cost to deviations from this position, the posture adopted to reach a point in space minimizes this cost [4], [5]. Movements are executed as a compromise between simultaneous, smooth interpolation of joint angles, minimizing discomfort, and straight hand paths. Rossetti *et al*. studied variability in pointing movements and found that errors increased at extreme joint positions [6]. They also concluded that configurations are chosen to minimize the sum of discomfort at the participating joints.

The impedance characteristics of the arm must be affected by kinematic properties of the human arm, motor control signals from the central nervous system (CNS), individual properties of each muscle, and proprioceptive feedback via the muscle spindle and Golgi tendon organ. For multi-joint hand movements, the hand stiffness and viscosity can be predicted with sufficient accuracy under the assumptions that the length of the muscle moment arm and the muscle viscoelasticity can be approximated by polynomial models of the joint angles [7]. Flash and Mussa-Ivaldi examined to what extent the kinematic properties of the human arm can explain its spatial variations and found that the anatomical factors are not sufficient to account for the observations [8].

Several studies have been made for single-joint and two-joint arm movements, where one human moved or/and regulated a task. Dowben has shown that the viscoelastic properties of skeletal muscles, which are the major source of human hand viscoelasticity, largely change depending on their activation level [9]. It has been also shown that the change of viscoelastic coefficients depends on the activation level of muscle [10], task instruction of the subjects [11], joint angles [12], and speed of the arm movement and loading [13]. Tsuji *et al*. pointed out that muscle contraction for a grip force increases stiffness and viscosity of the hand [14]. Also, Gomi *et al*. estimated hand stiffness during two-joint arm movements and argued that dynamic stiffness differs from static one because of the neuromuscular activity during movements [15], [16]. Tsuji analyzed the spatial characteristics of the human hand impedance with considering of effect of arm posture and muscle activity [7]. Gomi and Osu again showed that the stiffness and viscoelasticity of human multi-joint arm change under different contraction conditions during posture maintenance tasks and during force regulation tasks [17].

All the studies described are related to active and reaching movements. But it has been pointed out that passive movement is important for the cooperative task [18] and a variable structure of impedance characteristics is regulated by a motor command from the CNS. No attempts have been made to find out the characteristics of the human arm in passive movement and the time-variant nature of the impedance characteristic of the human arm.

An investigation has already been made into the impedance characteristics of the human arm's passive movements (the arm is moved by an external force) in the forward and backward direction while the forearm was in the horizontal position. Both the upper arm and forearm were in the same vertical plane. The elbow and the shoulder joint were assumed to have a constant center of rotation. The forearm was treated as a rigid body. In that investigation, mass, stiffness, and damping factor for the variable impedance model had been considered. It was found that the stiffness and the damping factor varied with the operational time [18].

In the present investigation, one degree-of-freedom rotational passive movements of the forearm around the elbow were considered. Both the upper arm and forearm were in the same horizontal plane. As only two muscles, biceps brachii and triceps brachii, are used in this rotational operation, the mechanics of the muscles and bones are simple and it is easier to analyze the characteristics of the musculoskeletal system [19]. To learn more about human motor adaptation, works have investigated the adaptation to stable [20]–[22] and unstable [23]–[25] interactions produced by a haptic interface.

In a cooperation task performed by two humans, one human control the position of the carried object and the other human follows the motion of that object. The former can designate as a leader and the latter as a follower. The characteristics of the follower can be applied to the control method of a cooperative robot. Moreover, if the target trajectory controlled by the leader is known, then the characteristics of the follower can be investigated easily as a simple spring-mass-damper system [1]. The arm of the human is moved along the target position trajectory and the force exerted by the arm is measured. From the data of the target position trajectory and force, the impedance characteristics of the human arm can be estimated.

The time trajectory of position and velocity found during the experiment of cooperation between two humans is similar to the minimum jerk motion proposed by Flash and Hogan [26] and Ikeura and Mizutani [27]. They found that the human arm moves to minimize (1) and (2).

$$J = \int_0^{t_f} \left(\frac{d^3\theta}{dt^3}\right)^2 dt \quad (1)$$

where t_f is the time duration of motion. The trajectory was derived by minimizing the function J as:

$$\theta(t) = \theta(0) + a\{10(t/d)^3 - 15(t/d)^4 + 6(t/d)^5\} \quad (2)$$

where a is movement amplitude, $\theta(0)$ is the position at time t_0 and d is the duration. The position and velocity for movement of 60^0 are shown in Figure 1.

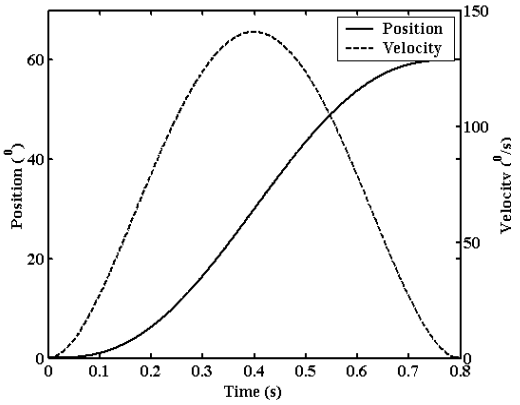


Figure 1. Time trajectory of position and velocity

The minimum jerk trajectory represents the free arm motion. Ikeura *et al*. found the minimum jerk trajectory in the cooperative motion of two humans [28]. This means that the tracking of the motion of the follower's arm is along the minimum jerk trajectory. In the cooperation between a human and a robot, the robot should follow the motion of the human so that the human can move his/her arm along the minimum jerk trajectory.

2. METHOD

2.1. Experimental set-up

Figure 2 illustrates an experimental system, in which a servo motor was used as the actuator. The servo motor was fixed into a frame vertically upward. One end of a splint (50×8 cm thin aluminum plate) was attached to the shaft of the motor. The arm was rotated along with the splint. A sensor located in between the arm and the splint was used to measure the torque needed to move the arm.

The output of the torque sensor was sent to the personal computer (PC) through the digital signal processing (DSP) board. An encoder was used to measure the angular position and the data was passed to the computer through the counter board. All boards were implemented on an industry standard architecture (ISA) bus of the PC.

2.2. Experimental procedure

The subjects are three right-handed male post-graduate university students (30-35 years old) with no previous history of neuropathies or trauma to the upper limbs. The subjects were given sufficient information about the experiment and then taken their consent to participate. In the experiment we defined a leader and a follower, the leader controls the position of the object and the follower tracks the motion of the object. Here, the robot was considered the leader and moved the splinter in the clockwise/anticlockwise directions. The leader controls the position, so the role of the robot was the same as a human leader at that time. The reason for choosing the robot as the leader was to move the arm at defined operating conditions.

As shown in Figure 2, a subject who followed the movement of the linear motor was seated beside the setup. The shoulder of the subject was restrained to the chair back and the elbow of the right arm was supported in the horizontal plane by a belt attached to the ceiling. He placed his arm on the splint so that the wrist was fitted into the torque sensor attached to the splint. Then the torque sensor was adjusted so that the elbow was positioned just above the center of rotation of the splint. A gap was maintained between the arm and splint as shown in Figure 3.

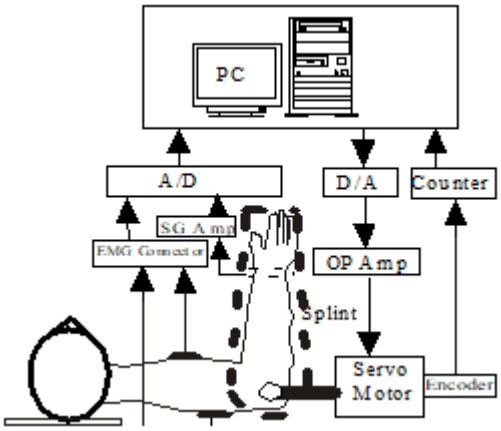


Figure 2. Experimental set-up

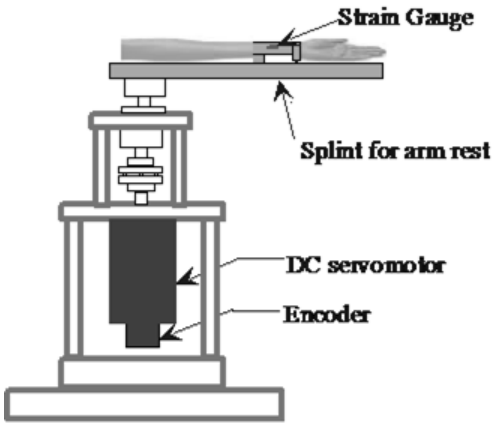


Figure 3. Servo-motor and splint

Figure 4 shows the control system of the experimental setup. Position trajectories used in the experiments were minimum jerk trajectories. The velocity trajectory was the first-order differentiation of the position trajectory. Movement amplitudes were 40^0-70^0 and the duration of the movements was from 0.6 to 1.2 seconds, with an increment of 0.2 seconds. The sampling time for the position control of the servo motor was 5 ms. The selection of position trajectory was done randomly so that the subject could not imagine the direction of rotation.

2.3. Data analysis

As the muscle is mechanically analogous to a spring-damper system, as shown in Figure 5, a simple second-order equation was used as the model for the arm dynamics. In the model, mass, damping factor, and stiffness were considered.

$$I_m\ddot{\theta} + c_m\dot{\theta} + k_m\theta = \tau \quad (3)$$

where I_m, c_m, and k_m are the impedance parameters for inertia, damping factor, and stiffness and τ is the torque to rotate the arm.

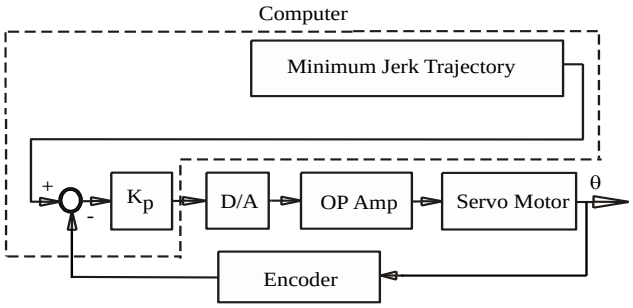


Figure 4. Block diagram of control system

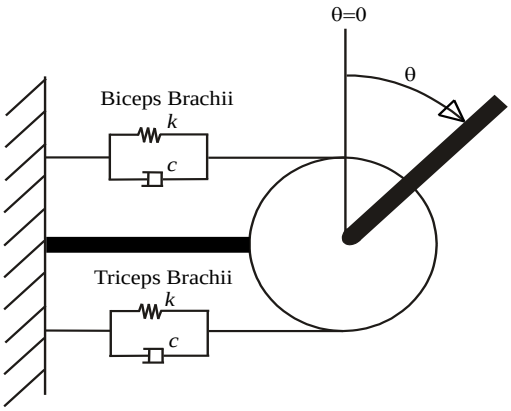


Figure 5. Impedance model of the human arm

For the estimation of the impedance parameters, the system identification toolbox of MATLAB (The Math Works, Inc.) was used [29]. For calculations, auto regressive exogenous (ARX) model was used. To make similarity with the ARX model, position $\theta(t)$ as an input and torque $\tau(t)$ as output was considered. If T is the sampling time then, $\dot{\theta}(t) = \frac{\theta(t)-\theta(t-1)}{T}$ and $\ddot{\theta}(t) = \frac{\dot{\theta}(t)-\dot{\theta}(t-1)}{T}$. By using these values in (3), obtained is (4).

$$\tau(t) = a_1\theta(t) + a_2\theta(t-1) + a_3\theta(t-2) \quad (4)$$

where, $a_1 = \frac{I+cT+kT^2}{T^2}$, $a_2 = \frac{-(2I+cT)}{T^2}$ and $a_3 = \frac{I}{T^2}$. In (4) is a form of the ARX model. Coefficients a_1, a_2, and a_3 were estimated by using the different variants of the recursive least-squares method. Then, the impedance parameters *I*, *c* and *k* were calculated.

3. RESULT

Figure 6 shows a typical time trajectory of position and torque measured during the experiments. This data was used for calculating the impedance parameters. Fifty-four replications were observed for the calculation of impedance parameters at different angles and speeds of movement. The angle of movement varied from 40 degrees to 70 degrees and the duration of movement varied from 0.6 seconds to 1.2 seconds with an interval of 0.2 seconds. Calculated impedance parameters of two operations are shown in Figure 7. Figure 7(a) represents the impedance parameter for the movement of 40 degrees in 0.6 seconds. A sample of impedance parameters for the movement of 70 degrees in 1 second is shown in Figure 7(b).

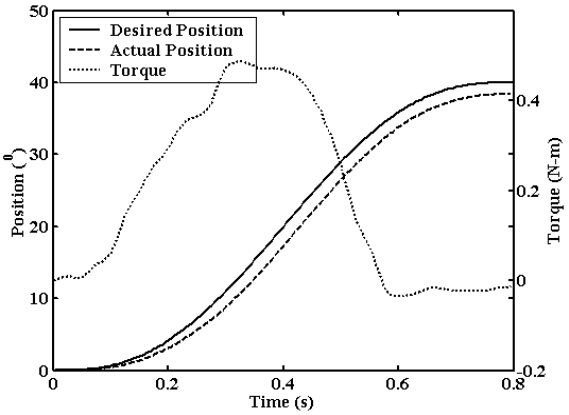
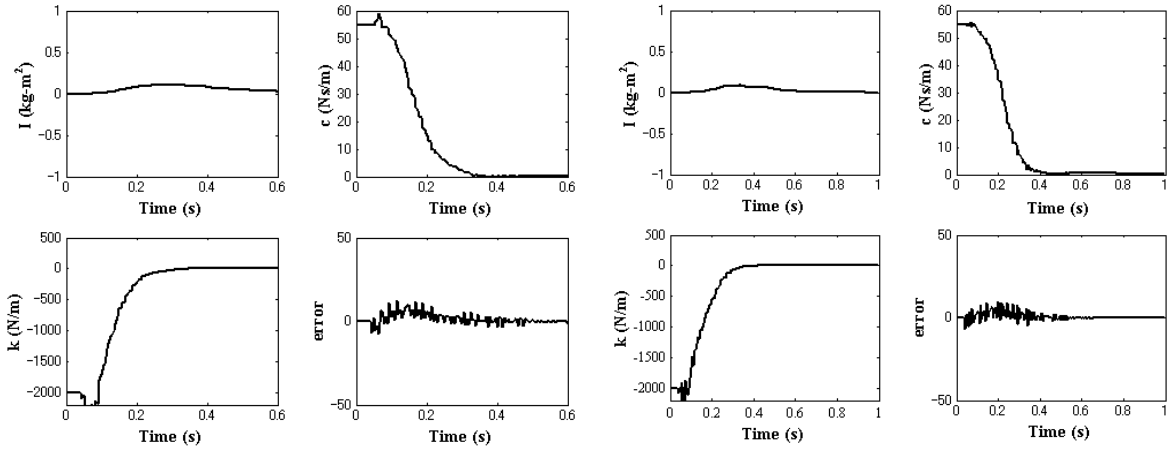


Figure 6. Time trajectories of the position and the torque



(a) (b)

Figure 7. Impedance parameters (a) a=40^0 and d=0.6 seconds and (b) a=70^0 and 1.0 second

4. DISCUSSION

In the present paper, the impedance of the human arm including inertia, stiffness, and damping factor was estimated for a single joint while it was moved by a robot. Figure 7 shows that the inertia is almost constant and the damping factor is high at the starting position and is near zero at 0.4 seconds. Stiffness has

also a similar characteristic to the damping factor. Similar results were found for the multi-joint arm movements with a higher viscous effect [1]. For a faster movement (Figure 7(a), a=40^0 and d=0.6 second) the parameters come to zero earlier (about 4% of total operation time). But in the case of other movements, the parameters come to zero at 0.4 seconds. Even for a very slow movement (a=40^0 and d=1.2 seconds) parameters come to zero at 0.4 seconds. Therefore, it is proved that the impedance characteristics of a human arm in passive movements do not depend upon the speed of movement or movement amplitude.

5. CONCLUSIONS

Impedance characteristics of the human arm during passive movement were analyzed. It is found that the impedance characteristics of a human arm for passive movements, maintain a model, which does not depend upon the speed and the movement amplitude, the inertia was constant, and the stiffness and damping factor varied from high to low within 0.4 seconds. Several subjects were used, and similar results were found.

ACKNOWLEDGEMENTS

Communication of this research is made possible through monetary assistance by Universiti Tun Hussein Onn Malaysia and the UTHM Publisher's Office via Publication Fund E15216.

REFERENCES

[1] M. Rahman, R. Ikeura, and K. Mizutani, "Cooperation characteristics of two humans in moving an object," *Machine Intelligence & Robotic Control*, vol. 4, no. 2, pp. 43–48, 2002.
[2] T. R. Kaminski and A. M. Gentile, "A kinematic comparison of single and multijoint pointing movements," *Experimental Brain Research*, vol. 78, no. 3, pp. 547–556, Dec. 1989, doi: 10.1007/BF00230242.
[3] H. Cruse and M. Brüwer, "The human arm as a redundant manipulator: the control of path and joint angles," *Biological Cybernetics*, vol. 57, no. 1–2, pp. 137–144, Aug. 1987, doi: 10.1007/BF00318723.
[4] H. Cruse, E. Wischmeyer, M. Brüwer, P. Brockfeld, and A. Dress, "On the cost functions for the control of the human arm movement," *Biological Cybernetics*, vol. 62, no. 6, pp. 519–528, Apr. 1990, doi: 10.1007/BF00205114.
[5] H. Cruse, M. Brüwer, and J. Dean, "Control of three- and four-joint arm movement: strategies for a manipulator with redundant degrees of freedom," *Journal of Motor Behavior*, vol. 25, no. 3, pp. 131–139, Sep. 1993, doi: 10.1080/00222895.1993.9942044.
[6] Y. Rossetti, C. Meckler, and C. Prablanc, "Is there an optimal arm posture? Deterioration of finger localization precision and comfort sensation in extreme arm-joint postures," *Experimental Brain Research*, vol. 99, no. 1, pp. 131–136, May 1994, doi: 10.1007/BF00241417.
[7] T. Tsuji, "Human arm impedance in multi-joint movements," *Advances in Psychology*, vol. 119, no. C, pp. 357–381, 1997, doi: 10.1016/S0166-4115(97)80013-1.
[8] T. Flash and F. Mussa-Ivaldi, "Human arm stiffness characteristics during the maintenance of posture," *Experimental Brain Research*, vol. 82, no. 2, pp. 315–326, Oct. 1990, doi: 10.1007/BF00231251.
[9] R. M. Dowben, "Contractility: medical physiology," *Mountcastle, VB and Mosby*, vol. C, no. 93, 1980.
[10] S. C. Cannon and G. I. Zahalak, "The mechanical behavior of active human skeletal muscle in small oscillations," *Journal of Biomechanics*, vol. 15, no. 2, pp. 111–121, Jan. 1982, doi: 10.1016/0021-9290(82)90043-4.
[11] F. Lacquaniti, F. Licata, and J. F. Soechting, "The mechanical behavior of the human forearm in response to transient perturbations," *Biological Cybernetics*, vol. 44, no. 1, pp. 35–46, May 1982, doi: 10.1007/BF00353954.
[12] W. A. MacKay, D. J. Crammond, H. C. Kwan, and J. T. Murphy, "Measurements of human forearm viscoelasticity," *Journal of Biomechanics*, vol. 19, no. 3, pp. 231–238, Jan. 1986, doi: 10.1016/0021-9290(86)90155-7.
[13] T. E. Milner, "Dependence of elbow viscoelastic behavior on speed and loading in voluntary movements," *Experimental Brain Research*, vol. 93, no. 1, pp. 177–180, Feb. 1993, doi: 10.1007/BF00227793.
[14] T. Tsuji, K. Goto, M. Morjtani, M. Kaneko, and P. Morasso, "Spatial characteristics of human hand impedance in multi-joint arm movements," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'94)*, 1994, vol. 1, pp. 423–430, doi: 10.1109/IROS.1994.407441.
[15] H. Gomi, Y. Koike, and M. Kawato, "Human hand stiffness during discrete point-to-point multi-joint movement," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1992, pp. 1628–1629, doi: 10.1109/IEMBS.1992.589839.
[16] H. Gomi and M. Kawato, "Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement," *Science*, vol. 272, no. 5258, pp. 117–120, Apr. 1996, doi: 10.1126/science.272.5258.117.
[17] H. Gomi and R. Osu, "Task-dependent viscoelasticity of human multijoint arm and its spatial characteristics for interaction with environments," *Journal of Neuroscience*, vol. 18, no. 21, pp. 8965–8978, Nov. 1998, doi: 10.1523/jneurosci.18-21-08965.1998.
[18] M. M. Rahman, R. Ikeura, and K. Mizutani, "Investigation of the impedance characteristic of human arm for development of robots to cooperate with humans," *JSME International Journal. Series C: Mechanical Systems, Machine Elements and Manufacturing*, vol. 45, no. 2, pp. 510–518, 2002, doi: 10.1299/jsmec.45.510.
[19] H. Kobayashi, R. Ikeura, and H. Inooka, "Evaluating the maneuverability of a control stick using electromyography," *Biological Cybernetics*, vol. 75, no. 1, pp. 11–18, Jul. 1996, doi: 10.1007/BF00238735.
[20] M. A. Conditt, F. Gandolfo, and F. A. Mussa-Ivaldi, "The motor system does not learn the dynamics of the arm by rote memorization of past experience," *Journal of Neurophysiology*, vol. 78, no. 1, pp. 554–560, Jul. 1997, doi: 10.1152/jn.1997.78.1.554.
[21] R. Shadmehr and H. H. Holcomb, "Neural correlates of motor memory consolidation," *Science*, vol. 277, no. 5327, pp. 821–825, Aug. 1997, doi: 10.1126/science.277.5327.821.
[22] S. N. Z. Ahmmad *et al.*, "Objective assessment of surgeon's psychomotor skill using virtual reality module," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1533–1543, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1533-1543.

[23] D. W. Franklin, R. Osu, E. Burdet, M. Kawato, and T. E. Milner, "Adaptation to stable and unstable dynamics achieved by combined impedance control and inverse dynamics model," *Journal of Neurophysiology*, vol. 90, no. 5, pp. 3270–3282, Nov. 2003, doi: 10.1152/jn.01112.2002.
[24] K. P. Tee, E. Burdet, C. M. Chew, and T. E. Milner, "A model of force and impedance in human arm movements," *Biological Cybernetics*, vol. 90, no. 5, pp. 368–375, May 2004, doi: 10.1007/s00422-004-0484-4.
[25] S. K. Debnath, R. Omar, and N. B. Abdul Latip, "Comparison of different configuration space representations for path planning under combinatorial method," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, p. 1, Apr. 2019, doi: 10.11591/ijeecs.v14.i1.pp1-8.
[26] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *Journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, Jul. 1985, doi: 10.1523/jneurosci.05-07-01688.1985.
[27] R. Ikeura and K. Mizutani, "Control of robot cooperating with human motion," in *Proceedings of 1998 IEEE International Workshop on Robotics and Human Communication*, 1998, pp. 525–529.
[28] R. Ikeura, H. Monden, and H. Inooka, "Cooperative motion control of a robot and a human," in *Robot and Human Communication - Proceedings of the IEEE International Workshop*, 1994, pp. 112–117, doi: 10.1109/roman.1994.365946.
[29] L. Ljung, *System identification toolbox, user's guide*. Natica, MA, USA: The Math Works, Inc., 1995.

BIOGRAPHIES OF AUTHORS

Md. Mozasser Rahman currently is an Associate Professor in the Department of Mechanical Engineering Technology, Universiti Tun Hussein Onn Malaysia (UTHM). Dr. Mozasser received a B. Sc. Eng. degree from Bangladesh Institute of Technology (BIT) Khulna in Mechanical Engineering in 1988. After graduation, he worked for the same institute as a lecturer. He got practical knowledge and experience in industrial maintenance and automation. He was later conferred an M. Eng. degree and Ph. D. from Mie University, Japan, in 2000 and 2003 respectively. Dr. Mozasser has expertise in robotics and industrial automation. His research area covers human-robot cooperation, movement characteristics of the human arm, and artificial human organsHe serves as a consultant for Industrial Automation, and Robotic Systems to universities and industries. He received 1 academic award from JSME (Japan Society of Mechanical Engineers) and 2 innovation awards from MTEX (Malaysian Technology Exhibition). One of his inventions is pending for patent. He published more than 50 articles and book chapters. He is a Member of the Institution of Mechanical Engineers, UK, and a Chartered Engineer registered with the Engineering Council, UK. He can be contacted at email: mozasser@uthm.edu.my.

Ryojun Ikeura currently is a Professor in the Department of Mechanical Engineering, Faculty of Engineering, Mie University, Japan. He received his B.E., M.E., and Ph.D. degrees in mechanical engineering from Tohoku University, Sendai, Japan, in 1986, 1988, and 1991 respectively. He has published more than 250 journal articles, conference papers, and book chapters. His research area covers human-robot cooperation, movement characteristics of the human arm and artificial human organs. He can be contacted at email: ikeura@ss.mach.mie-u.ac.jp.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 41~47
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp41-47

A new approach to achieve the users' habitual opportunities on social media

Arif Ridho Lubis, Mahyuddin K. M. Nasution, Opim Salim Sitompul, Elviawaty Muisa Zamzami
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received Oct 26, 2021
Revised Jul 19, 2022
Accepted Aug 17, 2022

Keywords:

Approach
Habitual
Naïve Bayes
Probability

ABSTRACT

The data generated from social media is very large, while the use of data from social media has not been fully utilized to become new knowledge. One of the things that can become new knowledge is user habits on social media. Searching for user habits on Twitter by using user tweets can be done by using modeling, the use of modeling lies when the data has been preprocessed, and the ranking will then be checked in the dictionary, this is where the role of the model is carried out to get a chance that the words that have been ranked will perform check the word in the dictionary. The benefit of the model in general is to get an understanding of the mechanism in the problem so that it can predict events that will arise from a phenomenon which in this case is user habits. So that with the availability of this model, it can be a model in getting opportunities for user habits on Twitter social media.

Corresponding Author:

Arif Ridho Lubis
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara
Padang Bulan 202155 USU, Medan, Indonesia
Email: arifridho.l@students.usu.ac.id

1. INTRODUCTION

The use of data on social media is very diverse and growing rapidly [1], [2]. Currently, social media produces huge amounts of data every day [3], [4]. This very large data can be used for various purposes and depends on what purpose the data is used for [5]. Consciously or unconsciously, Twitter social media users every time they tweet produce a maximum of 140 characters of letters in it, where each character forms a word and then forms a sentence. Sentences formed from words have their own meaning [6], to get the habits of users on Twitter social media, it can be seen from the frequent repeated words used by users. Where as in real life, activities or words that are often done are habits, habits are closely related to words that are verbs. Searching by utilizing available data in the world of social media is growing rapidly, but each method or method used by developers and researchers is different depending on the purpose and each method they do has advantages and disadvantages [7], [8].

The word habit on social media means a lot [9], where the habits of users on social media become new knowledge that can be used for other purposes, such as for the industrial world and other communities. One of the word searches uses the string match [10], [11], by utilizing the match string to get the number of words that are repeated from the initial data where the ranking applies according to the repetition of a word and is matched on a dictionary that has been labelled for look for the top-ranking word according to the labelled word in the dictionary. To facilitate the search, it is necessary to develop a new approach using mathematics where mathematics is the basic science of the computer itself. By utilizing modelling that serves to get an understanding or clarity of the mechanism in the problem so that it can predict events that will arise from a phenomenon.

2. MATERIAL AND METHOD

2.1. Document

Every document on social media can be used for research purposes. A social media document has many interrelated contexts, including user-provided annotations containing information [12]. The annotation itself contains tags [13], time, place, title and others that are closely related to user posts. From the structured context into a social media document, it is used as a research source, especially text data to become a very valuable source of information [14]. Text is also the simplest type of representation of information. Text documents include text document classification, grouping, topic detection, and several other processes [15].

Usually, the document is symbolized by D, so if there are several documents it becomes D_1, D_2, D_3… D_n. The document itself consists of a series of sentences that are interconnected with words, the word is symbolized by w. The number of words that may be obtained from a sentence so that it is symbolized by w_1, w_2, w_3…. w_n.

2.2. String matching

String matching technique is a pattern search in natural language processing, text, image processing, pattern and speech recognition that are commonly used [16]. There will be terms that are often encountered in string matching, namely patterns and text. The string matching algorithm is used to match a text with other text [17], [18]. A simple example of string matching is: i) Pattern: Watch and ii) Text: I watch animated movies on TV.

2.3. Basic probability

It is a statistical experiment which produces only one of many possible outcomes [19], [20]. The set of the whole possible oucames, the sample space is symbolized withl **Ω**. This sample space is also known as the set of events. Usually, a result can be denoted by ω. For an event probability is defined by P(.). For example, an experiment about choosing a word from a text "The definition of statistical experimentation can be widely stated as a process". So here the sample space is:

$$\Omega = \{a, statistical, experiment, can, be, broadly, defined, as, a, process\} \quad (1)$$

The incident occurred when choosing a word with:

$$\omega_1 = a, \omega_2 = statistical, w_3 = experiment, \dots, \omega_{10} = process \quad (2)$$

The total word count of the text is 10, or it is also known as the cardinality of **Ω**. Then it can be defined that the probability of choosing a word, for example, “experiment”, is written $P(experiment) = \frac{1}{10}$ or it can be written as:

$$P = \frac{The\ number\ of\ occurrences\ of\ choosing\ the\ word\ experiment}{The\ total\ number\ of\ events\ that\ can\ occur} \quad (3)$$

While the probability of choosing the word “a”, P(a) is $\frac{2}{10}$, because the number of words “a” in the text is 2. The complement of **Ω** is the incident of **Ω** not occured and symbolized by Ω^c with the note that $P(\Omega) = 1 - P(\Omega^c)$. Moreover $P(D|R)$ defines the conditional distribution of D where R is known. Furthermore, ∅ represents the empty set, that is ∅= {}. Suppose D and R are two occurrences of the sample space **Ω**, finite with N elements, this can be expressed as a Venn diagram as shown in Figure 1. The combination of events between D and R, can be described by $D \cup R$, where either event D or R or both occur [21].

In case of $D \cap R$, the intersection of occurrence D and R [22]. Here, they are considered mutually exclusive when the matter of one event prevents another’s event as shown in Figure 2. If the object is in ellipse R, what is the probability that the object is also in D. In order to be in D, the object must also be in slice D ∩ R. Therefore, the probability is equal to the number of components in |D ∩ R|, split by those in R, i.e., |R|. Officially the pattern is as (4):

$$P(D|R) = \frac{|D \cap R|}{|R|} = \frac{|D \cap R|/|\Omega|}{|R|/|\Omega|} = \frac{P(D \cap R)}{P(R)} \quad (4)$$

2.4. Architecture research

To be directed and precise according to the rules, it is limited by general architectural research as shown in Figure 3. So, the discussion will be more focused on what the author wants to get. From the figure, the following steps are obtained:

i) Initial data from Twitter that is ready for use, where the data has been pre-processed previously [19], where pre-processing follows from the purpose of this study.
ii) The preprocessing data is then ranked by word [23], where the assumption is that the word with repeated frequency is a word that refers to the habits of Twitter social media users.
iii) The results of this ranking will be checked by the campus that has been prepared, normally checking is done using string matching.
iv) Here the research environment plays a role between word ranking activities and checking the dictionary, where a new model is made to get opportunities for user habits on Twitter social media.

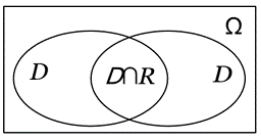


Figure 1. Sum of 2 events

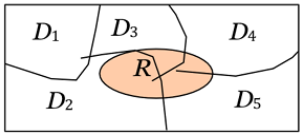


Figure 2. Mutually exclusive

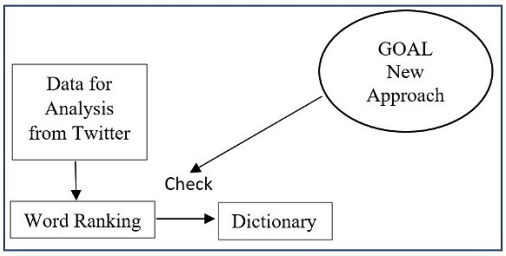


Figure 3. Architecture research

3. RESULTS AND DISCUSSION

3.1. Total probability theorem

The total probability theorem should be first comprehended before starting to discuss about Bayes' theorem, theorem [24]. Starting with the regulation to add two events *A* and *B*, it is clear form Figure 1 and Figure 2 that the (5). Suppose that the sample space is subdivided in n independent occurrence D_i, i=1.n, as seen in Figure 3. In Figure 3, it can be conluded if *DR* is expressed by $R = (R \cap D_1) \cup (R \cap D_2) \cup (R \cap D_3) \cup (R \cap D_4) \cup (R \cap D_5) \cup \cup (R \cap D_n)$ so that the total probability theorem can be obtained as:

$$P(D \cup R) = P(D) + P(R) - P(D \cap R) \quad (5)$$

$$= \sum_{i=1}^{n} P(R|D_i)P(D_i) \quad (6)$$

becomes

$$P(R) = P(R|D)\,P(D) + P(R|Dc)P(Dc) \quad (7)$$

as D2 ∪ D3 ∪ ... ∪ Dn is the complement of D1. Bayes' theorem [25] can be described:
Suppose $|D| \neq 0$ dan $|R| \neq 0$ The following conditions can be stated:

$$P(D|R) = \frac{|D \cap R|}{|R|} = \frac{P(D \cap R)}{P(R)} \quad (8)$$

$$P(R|D) = \frac{|R \cap D|}{|D|} = \frac{P(R \cap D)}{P(D)} \quad (9)$$

where the press (8) and (9) is clear as:

$$P(D \cap R) = P(D|R)P(R) = P(R|D)P(D) \quad (10)$$

as the result:

$$P(D|R) = \frac{P(R|D)P(D)}{P(R)} \quad (11)$$

When the sample **Ω** space can be split into a finite number of independent occurrences D_1, D_2, D_3, D_4, D_5 ...D_n, and when R constitute an occurrence with $P(R)>0$, that is the combined subset of all D_i, then for each D_i, then for each Di, Bayes formula which can be generalized as:

$$P(D_i|R) = \frac{P(R|D_i)P(D_i)}{\sum_{j=1}^{n} P(R|D_j)P(D_j)} \quad (12)$$

In (12) results from (11) due to the total probability theorem (6) and (7). With the recognized observational data, Bayes' theorem may be used to calculate the posterior probability of a hypothesis.

3.2. Naïve Bayes classification

Naive Bayesian learning adverts to the formation of a Bayesian probability model, which applies the posterior class of probability to a case: $P(Y=y_j|X = \mathbf{x}_i)$. Naive Bayes classifier [26] applies this probability to give a case to a class. Implement Bayes' Theorem (11) and simplify the notation, it would be obtained:

$$P(y_j|x_i) = \frac{P(x_i|y_j)P(y_j)}{P(x_i)} \quad (13)$$

Where the numerator in the (13) is the combined probability of $\mathbf{x}_i$ and y_j (10). As a result, the denominator can be changed into: only use **x**, omit the index *i* for simplify:

$$P(\mathbf{x}|yj)P\ (yj) = P(\mathbf{x}, yj) = P(x1|x2, x3, \ldots, xp, yj)\ P(x2|x3, x4, \ldots, xp, yj)\ P(xp|yj)\ P(yj)$$

Suppose the individual xi is not dependent one another. This is a strong hypothesis that clearly is against practical application, and thus Naive-as the suggested name. This supposition leads to $P(x_1|x_2, x_3, ..., x_p, y_j) = P\ (x_1|y_j)$. So, the combined probability of **x** and y_j is:

$$P(x|y_j) = \prod_{k=1}^{p} P(x_k|y_j)P(y_j) \quad (14)$$

which can be entered into press (13), so that there are:

$$P(y_j|x) = \frac{\prod_{k=1}^{p} P(x_k|y_j)P(y_j)}{P(x)} \quad (15)$$

It should be noted that the denominator, $P(\mathbf{x})$, has nothing to do with the category, for example for the categories y_j and y_l are the same. $P(\mathbf{x})$ works to be a scale factor and convinces that the posterior probability $P(y_j|\mathbf{x})$ is appropriately scaled. If we focus in clear classification rules, that is, to accurately assign each case to a class, we only need to calculate the numerator of each class and choose the maximum value of this value. The regulation is called the posterior maximum rule (16). The result class is also called the posterior maximum class (MAP), for the case of **x** it is calculated as yˆ:

$$y = argmax_{y_j} \widehat{\prod_{k=1}^{p}} P(x_k|y_j)P(y_j) \quad (16)$$

The maximum likelihood probability of a word belonging to a certain category is set by (17):

$$P(x_i|c) = \frac{Number\ of\ words\ x_i\ in\ class\ c\ document}{Total\ number\ of\ words\ in\ class\ c\ document} \quad (17)$$

According to Bayes' rule, the probability which a particular document corresponds to class c_i is given by:

$$P(c_i|d) = \frac{P(d|c_i)P(c_i)}{P(d)} \quad (18)$$

If we use the assumption of conditional free simple form, that we know a class, the words are conditionally independent of each other. Because of this assumption, the model is called Naïve.

$$P(c_i|d) = \frac{\prod P(x_i|c_j)P(c_j)}{P(d)} \quad (19)$$

Here xi is a word from the document. Classifier returns the class with the maximum posterior probability.

3.3. User

To determine what is the probability that the selected word is a verb. So, to determine the probability of the verb obtained in the dictionary, it is necessary to first describe the following things. For users, do the following steps to get a new model, as: i) number of users: U; ii) with U: $\{b_1, b_2, \dots, b_u\}$; iii) average number of texts user: T; iv) with $T = t^{b_1}, , t^{b_2}, \dots, t^{b_u}$; v) the meaning is $t^{b_1}$is the text from user b_1 , $t^{b_2}$is the text from user b_2; vi) average number of words per user: W; vii) with $W = w^{b_1}, w^{b_2}, \dots, w^{b_u}$; viii) average number of verbs from the dictionary: V; ix) so, for the whole community; x) the number of texts from all users is T.U; xi) the number of words for all users is W.U; xii) the order of occurrence of the word R; xiii) choose a word from the set W taken from the set of texts T from community U; xiv) the problem is to determine what is the probability that this selected word is a verb that belongs to the R occurrence ranking; and xv) the probability of choosing a word from the set W from the entire text is:

$$P(w^{b_i}|T) = \frac{WU}{TU} \quad (20)$$

Take the class t^{b_j} where j=1,2,…, u is the text class that comes from user $j \in U$. The probability that the class t^{b_j} given that the word w^{b_i} \ is in that class can be revised in conditional probability as:

$$P(t^{b_j}|w^{b_i}) = \frac{P(w^{b_i}|t^{b_j})P(t^{b_j})}{P(w^{b_i})} \quad (21)$$

In this equation $P(w^{b_i}|t^{b_j}) = \frac{WU}{TU}$ (from the Press. 20). The probability of choosing class t^{b_j} is $\frac{1}{T}$. While the probability of choosing the word w^{b_i} is $\frac{1}{W}$.

$$P(t^{b_j}|w^{b_i}) = \frac{(WU/TU)\,1/T}{1/W}$$

$$\text{or} = \frac{(WU/TU)\,W}{T}$$

WU is the total number of words from all users, TU total number of texts from all users, W total number (words)/user, T total number (texts)/user. Now we want to determine that the selected word is a verb. The number of verbs in the dictionary is V. The probability of choosing a word w^{b_i} knowing that it is a verb, is written as:

$$P(w^{b_i}|V) = \frac{WU}{V} \quad (22)$$

and the probability that the word w^{b_i} is contained in the V verb dictionary.

$$P(V|w^{b_i}) = \frac{P(w^{b_i}|V)P(V)}{P(w^{b_i})}$$

$$= \frac{(WU/V)\,1/V}{1/W}$$

$$= \frac{(WU)W}{V^2}$$

So now what we want to determine is the probability that the word w^{b_i} \ is included in the text of T and is included in the verb dictionary V. In other words, the word w^{b_i} is a verb.

$$P(w^{b_i}|V,T) = P(V|w^{b_i})P(T|w^{b_i})$$

$$P(V|w^{b_i})$$

Is the probability of selecting a verb from the user b_i. That is $\frac{w^{b_i U}}{v^2}$ W, $P(T|w^{b_i})$. It is the probability that the selected word from user b_i comes from the text. That is $\frac{w^{b_i}}{T}$ $\frac{w^{b_i}}{TU}$. With this new model, we can get a new approach to find opportunities for user habits on Twitter social media, making it easier for us to analyze user habits.

4. CONCLUSION

User habits on Twitter social media can be used for new knowledge. The same as getting a model by getting a mathematical model to look for word opportunities in the text that are included in the verb dictionary to be able to get users' habits on Twitter social media. Where is very helpful in terms of finding user habits. The design of the new model $\frac{w^{b_i}}{T}$ $\frac{w^{b_i}}{TU}$ can be used to find user habits. In the future, with this model, users can get habits, so that it is useful for other things such as in the field of product promotion, communities that have the same habits and others.

REFERENCES

[1] E. Amereihn *et al.*, "The social habit 2019." Edison Research, pp. 1–42, 2019.

[2] A. RA, R. S. Hegadi, and M. TN, "A big data security using data masking methods," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 7, no. 2, pp. 449–456, 2017, doi: 10.11591/ijeecs.

[3] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, "The effect of the TF-IDF algorithm in times series in forecasting word on social media," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, p. 976, 2021, doi: 10.11591/ijeecs.v22.i2.pp976-984.

[4] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, "Measuring information credibility in social media using combination of user profile and message content dimensions," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 3537–3549, 2020, doi: 10.11591/ijece.v10i4.pp3537-3549.

[5] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, "Obtaining value from the constraints in finding user habitual words," in *2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Oct. 2020, pp. 1–4, doi: 10.1109/ICADEIS49811.2020.9277443.

[6] M. AbdullahAl-Hagery, M. AbdullahAl-Assaf, and F. MohammadAl-Kharboush, "Exploration of the best performance method of emotions classification for arabic tweets," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 2, pp. 1010–1020, 2020, doi: 10.11591/ijeecs.v19.i2.pp1010-1020.

[7] G. U. Vasanthakumar, D. R. Shashikumar, and L. Suresh, "Profiling social media users, a content-based data mining technique for Twitter users," in *2019 1st International Conference on Advances in Information Technology (ICAIT)*, Jul. 2019, pp. 33–38, doi: 10.1109/ICAIT47043.2019.8987304.

[8] A. Mittal and S. Patidar, "Sentiment analysis on twitter data: a survey," in *ACM International Conference Proceeding Series*, 2019, pp. 91–95, doi: 10.1145/3348445.3348466.

[9] B. Gardner, A. L. Rebar, B. Gardner, and A. L. Rebar, "Habit formation and behavior change," *Psychology*, no. December, 2019, doi: 10.1093/obo/9780199828340-0232.

[10] D. E. Knuth, J. H. Morris, Jr., and V. R. Pratt, "Fast pattern matching in strings," *SIAM Journal on Computing*, vol. 6, no. 2, pp. 323–350, 1977, doi: 10.1137/0206024.

[11] W. Xuan, C. Ziyang, and W. Ding, "Uncertain string matching based on bitmap indexing," in *ACM International Conference Proceeding Series*, 2020, pp. 384–389, doi: 10.1145/3383972.3384007.

[12] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in *Proceedings of the third ACM international conference on Web search and data mining - WSDM '10*, 2010, p. 291, doi: 10.1145/1718487.1718524.

[13] H. Lian, Z. Qin, T. He, and B. Luo, "Knowledge graph construction based on judicial data with social media," in *2017 14th Web Information Systems and Applications Conference (WISA)*, Nov. 2017, pp. 225–227, doi: 10.1109/WISA.2017.46.

[14] Z. Wang, J. Liu, G. Sun, J. Zhao, Z. Ding, and X. Guan, "An ensemble classification algorithm for text data stream based on feature selection and topic model," in *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, Jun. 2020, pp. 1377–1380, doi: 10.1109/ICAICA50127.2020.9181903.

[15] W. H.Gomaa and A. A. Fahmy, "A survey of text similarity approaches," *International Journal of Computer Applications*, vol. 68, no. 13, pp. 13–18, 2013, doi: 10.5120/11638-7118.

[16] S. I. Hakak, A. Kamsin, P. Shivakumara, G. A. Gilkar, W. Z. Khan, and M. Imran, "Exact string matching algorithms: survey, issues, and future research directions," *IEEE Access*, vol. 7, pp. 69614–69637, 2019, doi: 10.1109/ACCESS.2019.2914071.

[17] I. Markic, M. Stula, and M. Zoric, "String pattern searching algorithm based on characters indices," 2019, doi: 10.23919/SpliTech.2019.8783109.

[18] M. AbuSafiya, "String matching algorithm based on letters' frequencies of occurrence," in *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, Jul. 2018, pp. 186–188, doi: 10.1109/CSIT.2018.8486384.

[19] G. Sarna and M. P. S. Bhatia, "A probalistic approach to automatically extract new words from social media," in *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 2016, pp. 719–725, doi: 10.1109/ASONAM.2016.7752316.
[20] L. V. Williams, *Probability, choice, and reason*. CRC Press, 2021.
[21] D. Berrar, "Bayes' theorem and Naive Bayes classifier," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.
[22] V. B. Nevzorov, M. Ahsanullah, and S. Annanjevskiy, *Probability theorry*. 2014.
[23] N. Ragavan, "Efficient key hash indexing scheme with page rank for category based search engine big data," *Proceedings of the 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2017*, vol. 2018-Febru, pp. 1–6, 2017, doi: 10.1109/ITCOSP.2017.8303118.
[24] A. S. Wagaman and R. P. Dobrow, *Probability: with applications and R*. Wiley, 2021.
[25] G. Michael and W. David, *Bayes linear statistics theory and methods*. Wiley, 2007.
[26] A. S. Chan, F. Fachrizal, and A. R. Lubis, "Outcome prediction using Naïve Bayes algorithm in the selection of role hero mobile legend," *Journal of Physics: Conference Series*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012041.

BIOGRAPHIES OF AUTHORS

Arif Ridho Lubis got master from Universiti Utara Malaysia in 2012 and graduate from Universiti Utara Malaysia in 2011, both information technology. He is a lecturer in Department of Computer Engineering and Informatics, Medan State Polytechnic in 2015. His research interest includes computer science, network, science and project management. He can be contacted at email: arifridho.l@students.usu.ac.id.

Mahyuddin K. M. Nasution is Professor from Sumatera Utara University, Medan Indonesia. Mahyuddin K. M. Nasution was born in the village of Teluk Pulai Dalam, Labuhan Batu Regency, North Sumatera Province. Worked as a Lecturer at the Sumatera Utara University, fields: mathematics, computer and information technology. Education: Drs. Mathematics (USU Medan, 1992); MIT, Computers and Information Technology (UKM Malaysia, 2003); Ph.D. in Information Science (Malaysian UKM). He can be contacted at email: mahyuddin@usu.ac.id.

Opim Salim Sitompul received the Ph.D. degree in information science from Universiti Kebangsaan Malaysia, Selangor, in 2005. He is currently a Professor with the Department of Information Technology, Sumatera Utara University, Medan, Indonesia. His skill and expertise are in AI, data warehousing, and data science. His recent projects are in natural language generation (NLP) and AIoT. The result of his work can be seen in his most recent publication is ''Template-Based Natural Language Generation in Interpreting Laboratory Blood Test.' He can be contacted at email: opim@usu.ac.id.

Elviawaty Muisa Zamzami graduated from Bandung Institute of Technology (Indonesia), magister of informatics, 2000 and awarded Doctoral in Computer Science from the University of Indonesia in 2013. She is a lecturer at Department of Computer Science, Sumatera Utara University, Indonesia. Currently her research interests are reverse engineering, requirements recovery, software engineering, requirements engineering and ontology. She can be contacted at email: elvi_zamzami@usu.ac.id.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 48~56
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp48-56

Improvement of transformer dissolved gas analysis interpretation using J48 decision tree model

Norazhar Abu Bakar[1], Imran Sutan Chairul[1], Sharin Ab Ghani[1], Mohd Shahril Ahmad Khiar[1], Mohd Zamri Che Wanik[2]
[1]High Voltage Engineering Research Laboratory, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, Kelantan, Malaysia

Article Info

Article history:

Received Sep 4, 2021
Revised Jul 26, 2022
Accepted Aug 24, 2022

Keywords:

Decision tree
Dissolved gas analysis
Doernenburg ratio method
J48
Transformer

ABSTRACT

Dissolved gas analysis (DGA) is widely accepted as an effective method to detect incipient faults within power transformers. Gases such as hydrogen, methane, acetylene, ethylene and ethane are normally utilized to identify the transformer fault conditions. Several techniques have been developed to interpret DGA results such as the key gas method, Doernenburg, Rogers, International Electro Technical Commission (IEC) ratio-based methods, Duval triangles, and the latest Duval pentagon methods. However, each of these approaches depends on the experts' shared knowledge and experience rather than quantitative scientific methods, therefore different diagnoses may be reported for the same oil sample. To overcome these shortcomings, this paper proposed the use of decision tree method to interpret the transformer health condition based on DGA results. The proposed decision tree model employed three main fault gases; methane, acetylene, ethylene as inputs, and classified the transformer into eight fault conditions. The J48 algorithm is used to train and developed the decision tree model. The performance of the proposed model is validated with the pre-known condition of transformers and compared with the Duval triangle method (DTM). Results show that the proposed model delivers better precision and accuracy in predicting transformer fault conditions compared to DTM with 81% and 69% respectively.

Corresponding Author:

Norazhar Abu Bakar
High Voltage Engineering Research Laboratory, Faculty of Electrical Engineering
Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Email: norazhar@utem.edu.my

1. INTRODUCTION

Dissolved gas analysis (DGA) has been used extensively to assess the power transformer condition. The decomposition of paper and oil occurs due to high thermal and electrical stress on the transformer insulation system, producing gases that dissolve in the oil and decrease its dielectric strength [1]. The subsequent paper decomposition yields carbon monoxide (CO) and carbon dioxide (CO_2). Hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), and ethane (C_2H_6), on the other hand, are produced as a result of oil decomposition and the formation of faults [2], [3]. Every fault produces unique characteristic gasses that can be used to identify the faults and measure their severity [4], [5]. The low energy level, partial discharge, produces H_2 and CH_4 gases while the high energy level, arcing, can produce all gases including C_2H_2. On the other hand, high thermal fault produces gases C_2H_4 and C_2H_6 [1]. Thus, the nature of the fault can be determined based on

the type of gas generates. However, the analysis is not always straightforward because at the same time there is a possibility of more than one fault occurred [6].

On the basis of the DGA findings, various interpretation methods such as key gas method (KGM), Doernenburg ratio method (DRM), Rogers ratio method (RRM), International Electro Technical Commission (IEC) ratio method (IRM), Duval triangle method (DTM) [7], and Duval pentagon method (DPM) [8] have been established to determine the transformer's state. However, most of the aforementioned approaches use the DGA test statistics used by professionals to provide knowledge-based diagnostic recommendations, while other approaches are based on the theory of thermodynamics, which may not necessarily lead to the same conclusion for the same oil sample [9]. Several soft computing methods were suggested to remove these limitations in order to overcome those problems.

To eradicate these shortcomings, artificial neural networks (ANN) [10]–[12], support vector machine (SVM) [13], [14] and fuzzy logic (FL) [15]–[17] were introduced. However, each of these methods also has some limitations. ANN is time-consuming, requires a large number of data samples to train the network properly to achieve consistent efficiency, requires a lot of time to learn and is prone to overfitting [18]. On the other hand, developing fuzzy rules and membership functions is tedious, and fuzzy outputs can be interpreted in a variety of ways that make analysis become complicated. In addition to being computationally expensive and complex, the key issue with SVM is the selection of the right function kernel [19]. Various kernel functions give different effects. This paper proposed another machine learning method, J48 decision tree to interpret DGA findings which offers a relatively quicker and less complex algorithm compared to SVM. Additionally, the structure of J48 decision tree is more comprehensible compared to ANN architecture.

2. DECISION TREE

Decision trees are one of the most effective methods in data mining for creating multiple covariates classification systems or for designing predictive algorithms for a target variable. This method is frequently used in numerous applications since it is user-friendly, straightforward, and stable even when missing values are present. In the decision tree method, a population will be classified into branch-like segments that create an inverted tree with a root node, internal nodes, and leaf nodes as shown in Figure 1 [20].

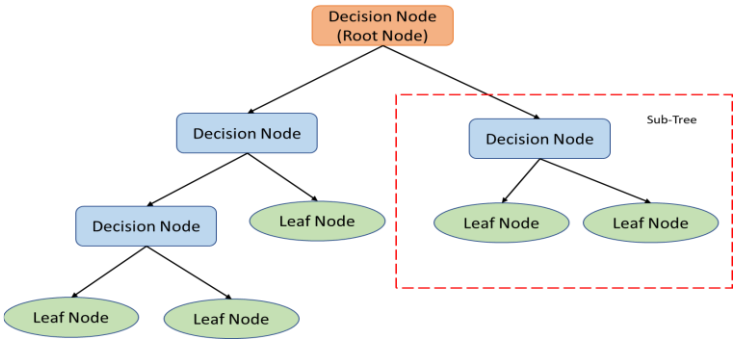


Figure 1. Decision tree model

A root node, also called a decision node represents a decision that will allow all records to be subdivided into two or more mutually exclusive subsets. The output of those decisions which do not contain any further branches is known as leaf nodes. Each leaf node symbolizes the mark of the particular class. On the other hand, several possible decisions available in the tree structure which connected between the root node and leaf nodes are called internal nodes.

Several decision tree algorithms like ID3, J48, CART, C5.0, SLIQ, SPRINT, random forest, and random tree have been developed for classification [21]. ID3, J48, and C5.0 algorithms implemented the top-down decision tree construction concept to obtain the output, while the CART algorithm is based on binary decision tree construction [22]. In this work, the J48 decision tree algorithm is chosen to classify the fault types of the transformer.

J48 decision tree or C4.5 algorithm developed by Ross Quinlan is an expansion of ID3 algorithm which allowed the target value of new test data to be decided with respect to the different attribute values of training data [20]. It improves the ID3 algorithm by dealing with both continuous and discrete attributes, missing values and pruning trees after construction. The J48 algorithm exploited a top-down greedy search through the given sets to test each attribute at every tree node [23].

As a supervised learning algorithm, a set of example data that consists of relationships between input objects and the desired output value is required to develop the J48 decision tree model [24]. This dataset will be used for training purposes. J48 decision tree induction methods begin with a root node representing the entire data set and separating the data into smaller subsets recursively by checking at each node for a given attribute. A root node is picked based on the highest gain values obtained among all attributes, while the splitting process is executed by considering the characteristics that are related to the degree of 'purity' in the dataset. This process is repeated until the subsets are "pure", whereas, all instances in the subset fall within the same class, at which time the tree growing is terminated. In the cases, where the stopping rules do not work well, then, the pruning process is conducted to decrease the classification errors [25]. Pruning is a process of removing the unnecessary nodes from a tree in order to get the optimal decision tree and also prevent the overfitting or underfitting rules been developed. The process of decision tree development using J48 algorithm is summarized in Figure 2.

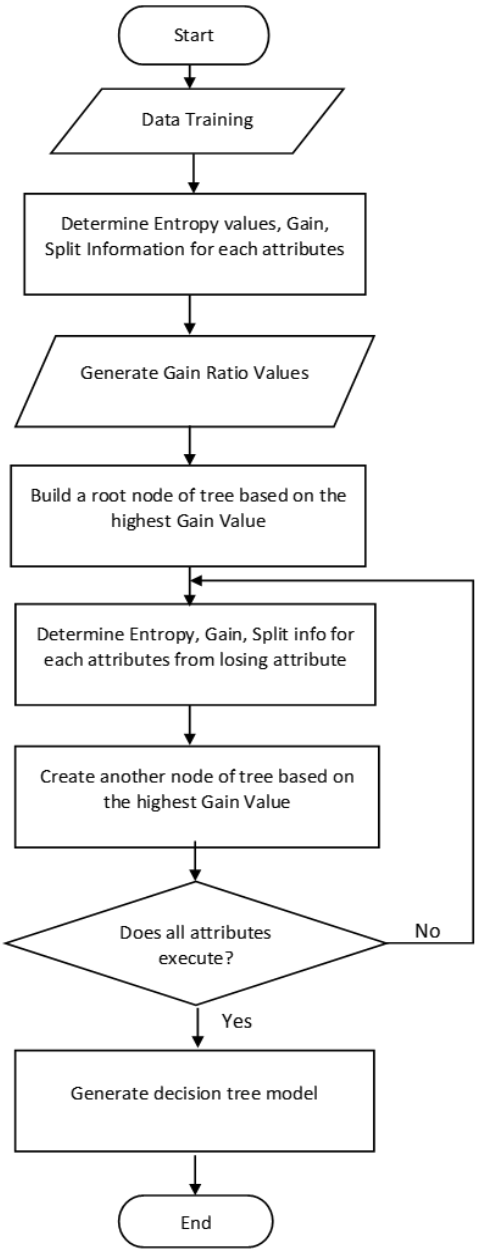


Figure 2. Flowchart of J48 (C4.5) algorithm

3. J48 DECISION TREE MODEL

To establish a DGA interpretation model, a total of 500 data collected from various operating transformers under different operating, age and health conditions were used to train by the J48 decision tree algorithm. Instead of using all fault gases, the proposed model only concentrated on the three main gases; CH_4, C_2H_2, and C_2H_4 as inputs attribute to interpret the transformer condition. These three gases are the same gases used in the DTM method. On the other hand, the output variable of the model that represents the transformer health conditions are classified into eight (8) categories as in Table 1.

The process of developing a DGA interpretation model is summarized in Figure 3. The process began by training a set of 500 transformers data with the known fault condition using J48 algorithm to obtain the decision tree model. These 500 datasets consist of all fault categories stated in Table 1. This training was performed using cross-validation with 10 folds procedure to increase the effectiveness of the proposed interpretation model.

Table 1. Proposed transformer fault classifications

Fault id	Fault description
NF	No fault
PD	Partial discharge
D1	Low energy discharge
D2	Arching
DT	Electrical-thermal
T1	Low thermal fault
T2	Medium thermal fault
T3	High thermal fault

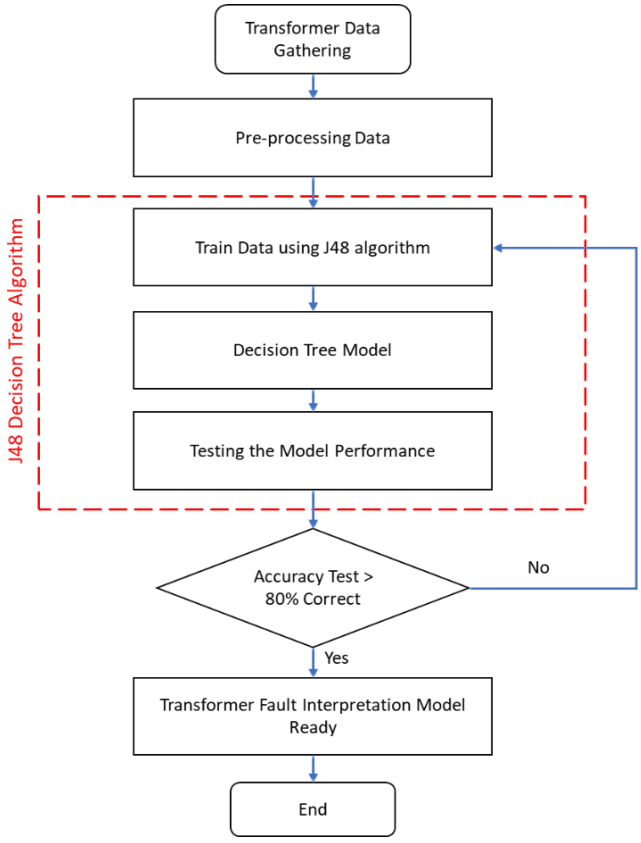


Figure 3. Process of developing the proposed interpretation model

The result of decision tree model generated with the J48 algorithm shows that the main attribute for the transformer faults among these three gases is C_2H_2, hence being selected as the root node in the model. The tree size of the developed model is 139 and consists of 70 leaves (leaf nodes). In the meantime, the developed decision tree model achieved 83.8% of correctness classified the transformer fault types whereas about 419 out of 500 datasets with 0.0619 and 0.1759 of mean absolute error (MAE) and root mean squared

error (RMSE) respectively. Detail prediction results are tabulated in the matrix as in Table 2. It can be seen that the proposed interpretation model developed was successfully classified each transformer fault type at an average of more than 80% except for D1, which a bit lower.

Table 2. Training results of J48 decision tree model

		J48 decision tree prediction							
		D1'	D2'	DT'	NF'	PD'	T1'	T2'	T3'
Actual	D1	32	16	0	1	1	1	0	0
	D2	3	138	0	4	1	1	1	2
	DT	0	1	8	1	0	0	0	0
	NF	0	0	1	39	0	2	0	0
	PD	1	2	0	1	30	3	0	0
	T1	3	0	0	3	3	93	1	7
	T2	1	1	0	0	0	3	13	4
	T3	0	2	2	0	0	7	1	66

After the best possible decision tree model has been achieved, an additional 100 datasets of known transformer fault types are used to evaluate further the performance of the proposed model. The proposed interpretation model must succeed at least 80% accuracy in classifying the overall transformer fault types before its ready to be used. Otherwise, the model will be modified and the training process is repeated until it succeeds 80% of prediction accuracy. From 100 datasets, the proposed model is able to correctly classified 81 of transformer faults as shown in Table 3, which equivalent to 81% of accuracy, hence surpassing the minimum requirement that has been agreed.

Table 3. Validation results of the proposed j48 decision tree model

		J48 decision tree prediction							
		D1'	D2'	DT'	NF'	PD'	T1'	T2'	T3'
Actual	D1	9	2	0	0	1	0	0	0
	D2	1	12	0	1	0	0	0	0
	DT	0	1	8	1	1	0	0	1
	NF	1	0	0	10	1	0	0	0
	PD	0	0	0	0	12	0	0	0
	T1	0	0	0	0	0	13	1	0
	T2	0	0	0	0	0	1	8	3
	T3	0	0	1	0	0	2	0	9

4. RESULTS AND DISCUSSION

In this section, the performance of the proposed interpretation model is compared with the DTM (as shown in Figure 4), which recognized as the best interpretation technique by industries so far. Although the latest improvement of DTM method is available, DPM, however its only works as a complementary to existing DTM, and does not replace it [8]. To evaluate the performance of both methods, another set of 65 transformers data with known fault conditions were used to examine the prediction accuracy. The confusion matrix is employed to analyze the performance of both methods in classifying the fault types. The confusion matrix is a table that reports the number of True positive (TP), True negative (TN), False positive (FP), and False negative (FN) which permits the visualization of classification accuracy and the performance of the method. The following are definitions of those terms:

i) TP: Cases in which correctly predicted Yes
ii) TN: Cases in which correctly predicted No
iii) FP: Cases in which predicted Yes, but actually is No
iv) FN: Cases in which predicted No, but actually is Yes.

The precision, recall, and F-measure are performed to examine the classification performance. The precision is to quantify the number of positive class predictions that actually belong to the positive class, while the recall will quantify the number of positive class predictions out of all positive examples in the dataset. On the other hand, F-measure provides a single score that balances both the concerns of precision and recall in one number. In the meantime, the accuracy of a classifier is referred to the probability of the method correctly predicting the actual fault of the transformer. The precision, recall, F-measure, and accuracy can be computed as:

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$recall = \left(\frac{TP}{TP+FN}\right) \quad (2)$$

$$F_{measure} = 2 \times \left(\frac{precision \times recall}{precision + recall}\right) \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

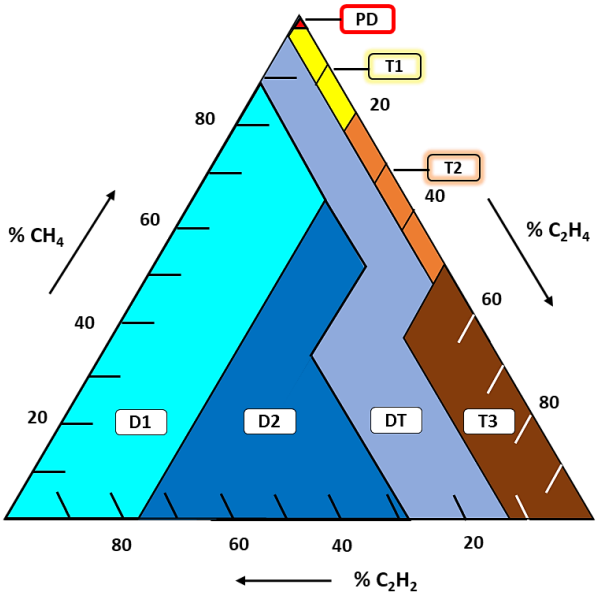


Figure 4. Duval triangle method diagram

Table 4 and Table 5 show the confusion matrix obtained for DTM and J48 Decision tree model respectively for 65 datasets of the transformer. According to Table 4, the DTM was successfully diagnosed 42 out of 65 cases, while the remaining cases were wrongly classified. On the other hand, the J48 model gives a better prediction with 53 out of 65 cases were correctly classified as shown in Table 5. Further analysis is shown in Table 6, whereas the precision, recall, F-Measure, and accuracy for each fault class are been analyzed.

Table 4. DTM confusion matrix

		J48 decision tree prediction							
		D1'	D2'	DT'	NF'	PD'	T1'	T2'	T3'
Actual	D1	8	0	0	0	0	0	2	0
	D2	3	8	3	0	0	0	0	1
	DT	0	0	2	0	0	0	1	0
	NF	0	0	0	0	0	1	2	1
	PD	0	0	2	0	1	0	0	1
	T1	0	0	1	0	1	6	1	2
	T2	0	0	0	0	0	0	6	0
	T3	0	0	0	0	0	0	1	11

Based on Table 4, it is noticed that the DTM is precisely classified the actual T2 fault (correctly classified 6 out of 6). However, it also frequently misinterprets other faults as T2, hence reducing the recall and accuracy of DTM in classifying T2 fault. In contrast with T1 results, although the DTM only manage to

correctly classified 6 out of 11 cases, however there is only 1 case where DTM is wrongly predicted. Therefore, the accuracy of DTM in classifying T1 fault is higher than T2. From results, it also noticed that the most truthfully classified by DTM is T3 with F-measure and accuracy are 0.79 and 0.65 respectively. The overall accuracy for DTM in classifying the fault types is only 40%.

On the other hand, the proposed J48 decision tree model has an average of 81% precisely classified fault types. The most precise class predicted by the J48 model is NF with 100% (4 out of 4) correct and followed by T3 with 92% (11 out of 12). Different from DTM, the J48 model generates more consistent interpretation results whereas the average recall achieved about 83%. The lowest recall is given by T2 whereas it is wrongly classified two cases as T2, which suppose to be T1 and T3. In the meantime, the proposed J48 model shows better accuracy compared to DTM with 69%.

Table 5. The proposed model confusion matrix

		J48 decision tree prediction							
		D1'	D2'	DT'	NF'	PD'	T1'	T2'	T3'
Actual	D1	8	1	0	0	1	0	0	0
	D2	2	12	0	1	0	0	0	0
	DT	0	0	2	0	0	0	0	1
	NF	0	0	0	4	0	0	0	0
	PD	1	0	0	0	3	0	0	0
	T1	0	1	0	0	0	8	1	1
	T2	0	1	0	0	0	0	5	0
	T3	0	0	0	0	0	0	1	11

Table 6. Precision, recall, f-measure and accuracy results comparison between DTM and the proposed J48 model

Fault types	Precision		Recall		F-measure		Accuracy	
	DTM	J48	DTM	J48	DTM	J48	DTM	J48
D1	0.80	0.80	0.73	0.73	0.76	0.76	0.62	0.62
D2	0.53	0.80	1.00	0.80	0.70	0.80	0.53	0.67
DT	0.67	0.67	0.25	1.00	0.36	0.80	0.22	0.67
NF	0.00	1.00	0.00	0.80	0.00	0.89	0.00	0.80
PD	0.25	0.75	0.50	0.75	0.33	0.75	0.20	0.60
T1	0.55	0.73	0.86	1.00	0.67	0.84	0.50	0.73
T2	1.00	0.83	0.46	0.71	0.63	0.77	0.46	0.63
T3	0.92	0.92	0.69	0.85	0.79	0.88	0.65	0.79
Average	0.59	0.81	0.56	0.83	0.53	0.81	0.40	0.69

5. CONCLUSION

This paper proposes a J48 decision tree model to interpret the transformer fault types based on the dissolved gas analysis data. The proposed model has been developed using a set of transformer historical data with the pre-known health condition. Three fault gases, CH_4, C_2H_4, and C_2H_2 are selected as inputs to the model and interpreted the transformer into eight fault classifications. The performance of the proposed model is evaluated using another sixty-five datasets and compared with the Duval Triangle method. Although the proposed model shows superior performance to DTM, however its accuracy can be improved further by considering more DGA samples during the training phase. Besides that, adding other fault gases such as H_2 and C_2H_6 also have the potential to enhance the model accuracy. However, by doing so, it may also increase the tree size and introduce overfitting issues if not considered carefully.

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Ministry of Higher Education Malaysia and Universiti Teknikal Malaysia Melaka for funding this study.

REFERENCES

[1] N. Bakar, A. Abu-Siada, and S. Islam, "A review of dissolved gas analysis measurement and interpretation techniques," *IEEE Electrical Insulation Magazine*, vol. 30, no. 3, pp. 39–49, 2014, doi: 10.1109/MEI.2014.6804740.

[2] IEEE, "C57.104-2008 - IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers," vol. 2008, no. February. pp. 1–45, 2008.

[3] IEEE, "C57.104-2019 - IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers." pp. 1–96, 2019.

[4] M. Duval and A. DePablo, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," *IEEE Electrical Insulation Magazine*, vol. 17, no. 2, pp. 31–41, 2001, doi: 10.1109/57.917529.
[5] H. C. Sun, Y. C. Huang, and C. M. Huang, "A review of dissolved gas analysis in power transformers," *Energy Procedia*, vol. 14, pp. 1220–1225, 2012, doi: 10.1016/j.egypro.2011.12.1079.
[6] A. Abu-Siada and S. Islam, "A new approach to identify power transformer criticality and asset management decision based on dissolved gas-in-oil analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 3, pp. 1007–1012, 2012, doi: 10.1109/TDEI.2012.6215106.
[7] M. Duval, "New techniques for dissolved gas-in-oil analysis," *IEEE Electrical Insulation Magazine*, vol. 19, no. 2, pp. 6–15, 2003, doi: 10.1109/MEI.2003.1192031.
[8] M. Duval and L. Lamarre, "The duval pentagon-a new complementary tool for the interpretation of dissolved gas analysis in transformers," *IEEE Electrical Insulation Magazine*, vol. 30, no. 6, pp. 9–12, 2014, doi: 10.1109/MEI.2014.6943428.
[9] A. Abu-Siada, "Improved consistent interpretation approach of fault type within power transformers using dissolved gas analysis and gene expression programming," *Energies*, vol. 12, no. 4, 2019, doi: 10.3390/en12040730.
[10] J. L. Guardado, J. L. Naredo, P. Moreno, and C. R. Fuerte, "A Comparative Study of Neural Network Efficiency in Power Transformers Diagnosis Using Dissolved Gas Analysis," *IEEE Power Engineering Review*, vol. 21, no. 7, p. 71, 2001, doi: 10.1109/MPER.2001.4311493.
[11] S. A. Khan, M. D. Equbal, and T. Islam, "A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 1, pp. 590–596, 2015, doi: 10.1109/TDEI.2014.004478.
[12] S. S. M. Ghoneim, I. B. M. Taha, and N. I. Elkalashy, "Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 3, pp. 1838–1845, 2016, doi: 10.1109/TDEI.2016.005301.
[13] Y. Zhang *et al.*, "A Fault Diagnosis Model of Power Transformers Based on Dissolved Gas Analysis Features Selection and Improved Krill Herd Algorithm Optimized Support Vector Machine," *IEEE Access*, vol. 7, pp. 102803–102811, 2019, doi: 10.1109/ACCESS.2019.2927018.
[14] J. Li, Q. Zhang, K. Wang, J. Wang, T. Zhou, and Y. Zhang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, pp. 1198–1206, 2016, doi: 10.1109/TDEI.2015.005277.
[15] A. Abu-Siada, S. Hmood, and S. Islam, "A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 6, pp. 2343–2349, 2013, doi: 10.1109/TDEI.2013.6678888.
[16] N. A. Bakar and A. Abu-Siada, "Fuzzy logic approach for transformer remnant life prediction and asset management decision," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 5, pp. 3199–3208, 2016, doi: 10.1109/TDEI.2016.7736886.
[17] I. B. M. Taha, A. Hoballah, and S. S. M. Ghoneim, "Optimal ratio limits of rogers' four-ratios and IEC 60599 code methods using particle swarm optimization fuzzy-logic approach," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, no. 1, pp. 222–230, 2020, doi: 10.1109/TDEI.2019.008395.
[18] M. E. A. Senoussaoui, M. Brahami, and I. Fofana, "Combining and comparing various machinelearning algorithms to improve dissolved gas analysis interpretation," *IET Generation, Transmission and Distribution*, vol. 12, no. 15, pp. 3673–3679, 2018, doi: 10.1049/iet-gtd.2018.0059.
[19] M. Zhang, K. Li, and H. Tian, "Multiple SVMs modelling method for fault diagnosis of power transformers," *Przeglad Elektrotechniczny*, vol. 88, no. 7 B, pp. 232–234, 2012.
[20] J. R. Quinlan, *{C4}.5 - Programs for Machine Learning*. Elsevier, 2014.
[21] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015, doi: 10.11919/j.issn.1002-0829.215044.
[22] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manufacturing*, vol. 35, pp. 698–703, 2019, doi: 10.1016/j.promfg.2019.06.011.
[23] C. Anuradha and T. Velmurugan, "A data mining based survey on student performance evaluation system," in *2014 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2014*, 2015, pp. 43–47, doi: 10.1109/ICCIC.2014.7238389.
[24] M. A. Muslim, A. Nurzahputra, and B. Prasetiyo, "Improving accuracy of C4.5 algorithm using split feature reduction model and bagging ensemble for credit card risk prediction," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Mar. 2018, vol. 2018-Janua, pp. 141–145, doi: 10.1109/ICOIACT.2018.8350753.
[25] R. Panigrahi and S. Borah, "Rank Allocation to J48 Group of Decision Tree Classifiers using Binary and Multiclass Intrusion Detection Datasets," *Procedia Computer Science*, vol. 132, pp. 323–332, 2018, doi: 10.1016/j.procs.2018.05.186.

BIOGRAPHIES OF AUTHORS

Norazhar Abu Bakar received his BEng. (Hons.) degree in Electronics and Electrical Engineering from Leeds University, UK, MSc. (Eng.) in Control Systems from Sheffield University, UK, and PhD degree in Electrical and Computer Engineering from Curtin University, Australia. He started his career as Assembly Engineer at Toshiba Electronics, Malaysia, and then moved to TNB as SCADA/DA Project Engineer. Currently, he is serving as Senior Lecturer at Faculty of Electrical Engineering, UTeM. He is a member of the High Voltage Engineering Research Laboratory in UTeM. His research interests are condition monitoring, asset management, and artificial intelligence. He can be contacted at email: norazhar@utem.edu.my.

Imran Sutan Chairul was born in Kuala Lumpur, Malaysia, in 1984. He received his BEng. (Hons.) degree in Electrical Engineering from UTeM in 2008 and MEng. degree in Electrical Engineering from National Energy University in 2012. He is currently pursuing his PhD degree at UTeM, where his research is focused on vegetable-based transformer dielectric liquids. He can be contacted at email: imransc@utem.edu.my.

Sharin Ab Ghani received his BEng. (Hons.) degree in Electrical Engineering from UTeM in 2008, MEng. Degree in Electrical Engineering from National Energy University in 2012, and PhD degree in Electrical Engineering from Malaysian University of Technology in 2019. Currently, he is serving as Senior Lecturer at Faculty of Electrical Engineering, Technical University of Malaysia Melaka, and Head of Energy and Power Systems (EPS) Research Group. His research interests are centered on high voltage engineering, power equipment condition monitoring, green electrical insulation, design of experiments, and optimization. To date, his outstanding research works have been published in WoS-SCI (3) and Scopus (29) indexed journals. He can be contacted at email: sharinag@utem.edu.my.

Mohd Shahril Ahmad Khiar was born in Selangor, Malaysia in 1984. He received his BSc. in Electrical & Electronics Engineering from Korea University in 2008, Master's degree in Electrical Engineering from National Energy University in 2012, and PhD degree from University of Southampton, UK in 2019. He is currently serving as Senior Lecturer at Faculty of Electrical Engineering, UTeM and Head of High Voltage Engineering Research Laboratory in UTeM, where he is also the Deputy Head of Energy and Power Systems (EPS) Research Group. He can be contacted at email: mohd.shahril@utem.edu.my.

Mohd Zamri Che Wanik received his BSc. from the University of Evansville, U.S.A., MEngSc. from the Curtin University of Technology, Australia and Doktor der Ingenieurwissenschaften (equivalent to PhD) from Universität Duisburg-Essen, Germany in 1997, 2002 and 2011 respectively all in Electrical Engineering specialising in Electrical Power System. Currently he is serving as a Visiting Professor at the Institute for Artificial Intelligence and Big Data, University of Malaysia Kelantan, City Campus, Pengkalan Chepa, 16100 Kota Bharu, Kelantan, Malaysia. He is a senior scientis at Qatar Environment and Energy Research Institute (QEERI), which is associated with Hamad Bin Khalifa University, Qatar in the capacity of project lead of Smart Grid. He can be contacted at email: mwanik@hbku.edu.qa

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 57~65
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp57-65

K-nearest neighbor based facial emotion recognition using effective features

Swapna Subudhiray, Hemanta Kumar Palo, Niva Das
Department of Electronics and Communication Engineering, Siksha 'o' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India

Article Info

Article history:

Received Oct 22, 2021
Revised Jul 23, 2022
Accepted Aug 21, 2022

Keywords:

Histogram of oriented gradients
K-nearest neighbor
Local binary pattern
Recognition of facial expression

ABSTRACT

In this paper, an experiment has been carried out based on a simple k-nearest neighbor (kNN) classifier to investigate the capabilities of three extracted facial features for the better recognition of facial emotions. The feature extraction techniques used are histogram of oriented gradient (HOG), Gabor, and local binary pattern (LBP). A comparison has been made using performance indices such as average recognition accuracy, overall recognition accuracy, precision, recall, kappa coefficient, and computation time. Two databases, i.e., Cohn-Kanade (CK+) and Japanese female facial expression (JAFFE) have been used here. Different training to testing data division ratios is explored to find out the best one from the performance point of view of the three extracted features, Gabor produced 94.8%, which is the best among all in terms of average accuracy though the computational time required is the highest. LBP showed 88.2% average accuracy with a computational time less than that of Gabor while HOG showed minimum average accuracy of 55.2% with the lowest computation time.

Corresponding Author:

Swapna Subudhiray
Department of Electronics and Communication Engineering, Siksha 'o' Anusandhan (Deemed to be University)
Bhubaneswar, Odisha, India
Email: swapnaray89@gmail.com

1. INTRODUCTION

Among many modalities of human affective states, the facial expression remains a significant mode of communicating an individual's state of mind. Facial expression accounts for 55% of the entire emotional information as compared to 38% by discourse, and 7% by language [1], [2]. Among these modalities, the recognition of emotions using facial expression (RFE) remains a complex domain of research due to the absence of standard best features adequately describing these states. It remains significant in the area of human-machine interaction and design acknowledgment [3]–[5].

There are two major approaches to facial emotion recognition as appearance and geometric-based model [6], [7]. However, the techniques based on geometric models do not consider the skin surface adjustments such as the significant wrinkles displaying the outward appearance. On the contrary, appearance-based techniques utilize the whole face or unequivocal zones in the facial image to represent the shrouded information [8]–[10].

In this regard, the Gabor Filter is an appropriate strategy to recognize human expressive states with promising results earlier. The technique is suitable for extracting information on multi-scale, multi-course changes in an expressive facial surface while not disturbing the changes in brightness. It targets the prominent features of emotion by focusing on the variation in the edge and texture of an image [11], [12]. On

the contrary, the histogram of oriented gradient (HOG) process develops the histogram corresponding to each cell comprising several pixels by estimating the luminance gradient of each pixel. It is a geometric-based approach in which, the luminance gradient utilizes all the adjacent pixels such as the top, bottom, left, and right, to compute the magnitude and the direction of the variation in color intensity of a cell. The main properties of local binary pattern (LBP) are obstruction against brilliance changes and their computational ease [13]. However, the applicability of the HOG and LBP technique in RFE as compared to the Gabor filter under different training and testing data, orientation and Kappa coefficients can provide new insights to researchers, thus investigated here [14].

Classification algorithms play an important role in the identification of facial expressions (FE). Earlier literature in RFE has explored several classification mechanisms such as random forest, naive Bayes, support vector machine (SVM), Hidden Markov model (HMM), AdaBoost, multilayer neural networks, decision tree, K-nearest neighbors, and deep neural networks with excellent results [13]–[15]. Nevertheless, reliable, comprehensive, and faster classification algorithms are often chosen which should address the challenges of subject-dependency, variation in illumination, and the position of the head during the affective states [16].

Here Section 2 investigates the chosen feature extraction techniques in detail. Section 3 briefs the choice of the database whereas the reason for choosing the k-nearest neighbor (kNN) classifier has been provided in section 4. The simulation results using the chosen classifier and the extracted feature sets have been explained in section 5 and lastly, section 6 concludes the work with future directions.

2. FEATURE EXTRACTION METHOD

The facial image identification modeling is shown in Figure 1. It comprises several components meant for image acquisition, pre-processing, feature extraction, and classification. After clicking an FE image using a camera, it is pre-processed to minimize any variation due to the environment and other sources. The pre-processing step involves image-scaling, adjustment of contrast and brightness, and image enhancement. As the facial images of the chosen Japanese female facial expression (JAFEE) and Cohn-Kanade (CK+) database have already been pre-processed, it is not required to involve this step here. This work explores the Gabor filter, LBP, and HOG. Feature extraction techniques to classify the FE states using facial images. The feature extraction techniques have been briefly explained in the following subsections.

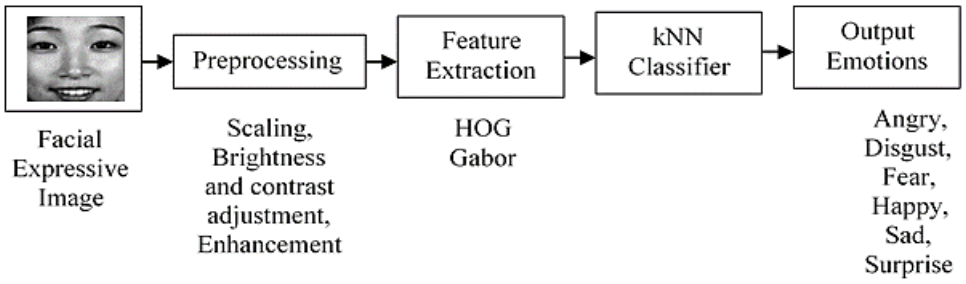


Figure 1. The facial image identification modeling

2.1. Histogram of oriented gradients

HOG technique is considered here as it focuses on both local and global facial expression attributes in different orientations and scales. The features are sensitive to variations in the shape of an object unless the shape is consistent [17]. This piece of work utilizes nine bin histograms representing the directions and strength of an edge using 4×4 cells corresponding to each patch. These features of each active facial patch are appended to extract the desired feature vector [18].

For the pixel $z(s,t)$, the gradient is computed in the HOG approaches,

$$G_p = z(s-1,t) - z(s+1,t) \quad (1)$$

$$G_q = z(s,t-1) - z(s,t+1) \quad (2)$$

The gradient magnitude is given by,

$$G = \sqrt{{G_p}^2 + {G_q}^2} \quad (3)$$

The orientation of the bin is given by,

$$\theta = arctan(G_q/G_p) \quad (4)$$

Where θ denotes the bin angle. Both the magnitude and the bin angle are used to form the HOG feature vector. In this work, the size of each HOG cell is fixed at 8×16 pixels. This way, it is possible to focus on the variation of the shape of the eyes, mouth, and eyebrows that change more vertically during an emotional outburst. To choose the cell size, we begin with 2×2 pixels to 64×64 pixels using all the possible variations in both vertical and horizontal dimensions and noting the RFE accuracy. The cell size of 8×16 has provided the highest accuracy, and hence is kept for further processing. It is observed that with an increase in cell size, there is a loss of image details, and the computation time increases. On the contrary, the feature vector dimension remains small and the computation time becomes faster with smaller-sized cells.

2.2. Local binary pattern

LBP is a very popular, efficient, and simple texture descriptor that is used for many computer vision problems [19]. It can capture the spatial pattern along with the grayscale contrast using a simple thresholding technique, where the intensities of the neighboring pixels are compared with that of the center pixel resulting in a binary pattern termed LBP [20]. The basic LBP operation with a 3×3 window is expressed and demonstrated.

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \quad (5)$$

Where i_c corresponds to the intensity of the central pixel (x_c, y_c), i_n corresponds to the gray values of the eight closed pixels, and if $i_n - i_c > 0$, then $s(i_n - i_c) = 1$, else $s(i_n - i_c) = 0$.
The mathematical form is donated as,

$$LBP_{P,R}^{U2} = \sum_{j=0}^{p-1} S(g_j - g_c)2^j \quad (6)$$

where the gray value of the jth pixel is g_j and the gray value of the ith pixel is g_c respectively, S(x) is a unit step function defined.

$$S(x) = \begin{cases} 1, & if(x \geq 0) \\ 0, & if(x < 0) \end{cases}$$

The multi-goal examination can be accomplished by picking various estimations of R and P. Figure 2 shows three diverse sweeps of LBP administrators. From left to right, they are $LBP_{4,1}^{U2}, LBP_{8,1}^{U2}$, and $LBP_{8,2}^{U2}$ operators respectively.

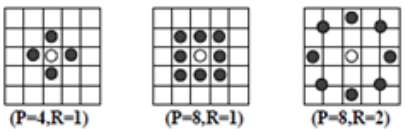


Figure 2. An example of a basic LBP operation

After applying the LBP operator to an image, the histogram is calculated,

$$H_i = \sum_{x,y} I(f_1(x, y) = 1), \mathrm{i} = 0, \ldots, \mathrm{n} - 1 \quad (7)$$

Here *n*= different labels and,

$$I(A) = \begin{cases} 1, A\ is\ True \\ 0, A\ is\ False \end{cases}$$

2.3. Gabor filters

Gabor filter is a linear filter and is described by the spatial and frequency domain representation of the signal. It can provide important information on emotions as the filter can approximate the human's perception adequately [21]. It can be expressed as a combination of the complex exponential function and the 2D Gaussian function.

$$f(a,b) = exp\left(-\frac{{a_1}^2+\gamma^2{b_1}^2}{2\sigma^2}\right) exp\left(j\left(2\pi\frac{a_1}{\lambda}+\varphi\right)\right) \quad (8)$$

Where $a_1 = acos\theta + bsin\theta$ and $b_1 = -asin\theta + bcos\theta$. Here $\theta, \lambda, \varphi, \sigma,$ and γ denotes the orientation in degrees, wavelength, phase offsets, standard deviation, and the spatial aspect ratio respectively. Using the real component of (8), the expression for the Gabor filter becomes,

$$f(a,b) = exp\left(-\frac{{a_1}^2+\gamma^2{b_1}^2}{2\sigma^2}\right) cos\left(2\pi\frac{a_1}{\lambda}+\varphi\right) \quad (9)$$

this work develops the Gabor filters using a 39×39 size pixel window. Earlier researchers in this direction have employed approximately seven or eight different values of θ and four to five different values of λ. However, for our purpose, three different values of $\lambda = \{3, 8, 13\}$ and four different values $\theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \pi\}$ have been chosen after a few iterations while keeping the parameters $\gamma = 0.5$, $\sigma = 0.56\lambda,$ and $\varphi = 0$ as constant [22]. The input image I is convolved with Gabor filter f to extract the Gabor features F for a specific θ and λ.

$$F_{\lambda,\theta} = I * f_{\lambda,\theta} \quad (10)$$

3. PROPOSED METHOD

3.1. Japanese female facial expression (JAFFE) database

The JAFFE database is easily accessible and has been chosen by several researchers in the RFE, which makes the comparison platform uniform, hence considered here. The images are stored on a grayscale with a resolution of 256×256. The happy, disgust, fear, angry, neutral, sad, and surprising emotional expression samples from the JAFFE database has been provided in Figure 3. We have considered 188 images consisting of six basic emotions in this work.

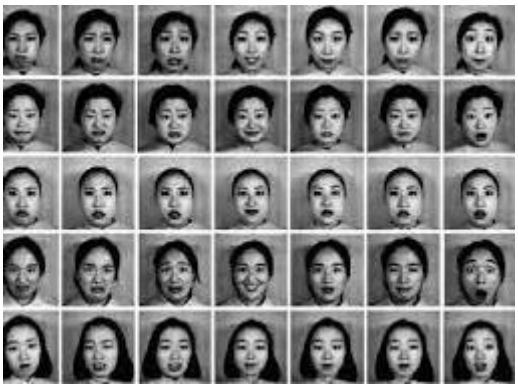


Figure 3. Sample images of the JAFFE database

3.2. CK+ database

The extended CK+ information base contains outward appearances of 123 college students. In the information base, we chose 928 picture groupings from 123 subjects, with 1 to 6 feelings for every subject. There are 928 images comprising 135 anger, 207 joy, 84 sad, 249 surprises, 75 fears, and 177-disgust FEs. Figure 4 provides the sample images of CK+ emotional expressive states.

3.3. Classification

kNN is a non-parametric supervised learning algorithm meant for classification as well as regression. It relies on the concept of feature similarity to classify new data meaning. In this, the new data

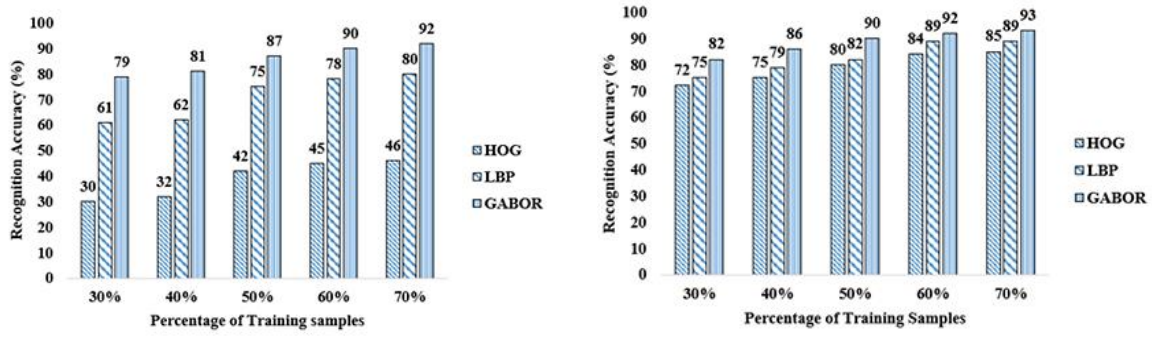
will be assigned a class based on how closely it matches the data in the training set [23]. It allocates the feature variable to the designated class based on a distance measure such as the Euclidean norm. For vectors $p = (p_1, p_2 \dots \dots \dots p_m)$ and $q = (q_{1,} q_2 \dots \dots \dots q_m)$, the distance norm can be expressed as,

$$d(p,q) = \sqrt{\sum_{j=1}^{m} (p_j - q_j)^2} \quad (11)$$

Figure 4. Sample images of CK+ emotional expressive states

4. RESULTS AND DISCUSSIONS

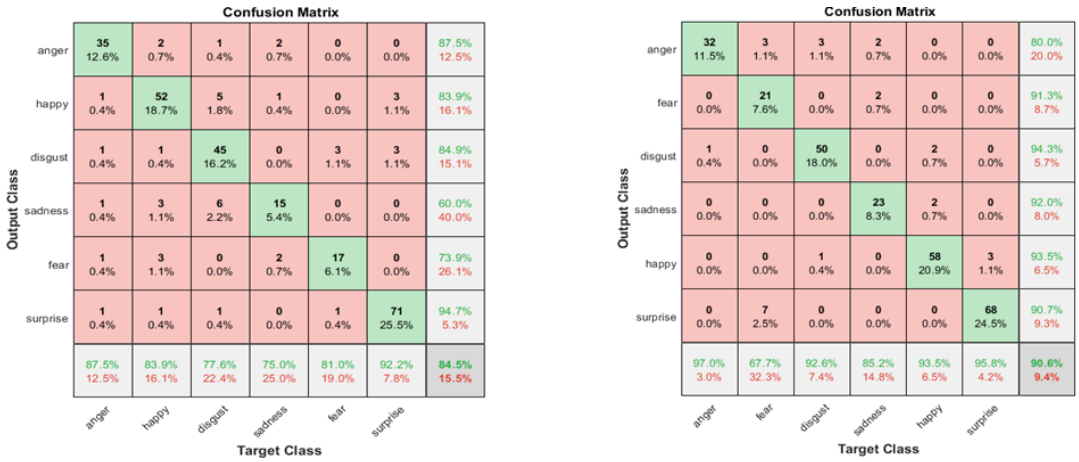
The kNN classifier has been utilized to order the extracted feature sets into six different basic emotions. Different training and testing data division ratios such as 70%/30%, 60%/40%, 50%/50%, 40%/60%, and 30%/70% have been trialed from the chosen JAFFE and CK+ database to access the best possible recognition accuracy with the classifier. A data division ratio of 70%/30% has provided the desired level of accuracy and hence retained for this work. Figure 5 compares kNN accuracy using the extracted feature sets with different data division ratios for JAFEE and CK+ Dataset. Figure 5 (a) shows the kNN accuracy for the JAFEE dataset whereas Figure 5 (b) shows the accuracy for the CK+ dataset.



(a) (b)

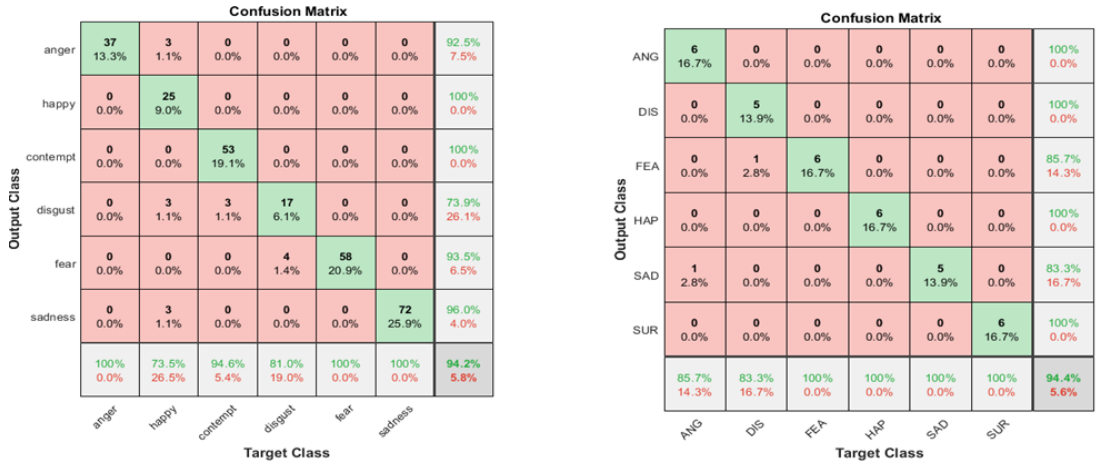
Figure 5. The comparison of kNN accuracy using the extracted feature sets with different data division ratios for, (a) JAFEE dataset and (b) CK+ dataset

Training and testing were carried out on three sets of features, i.e., HOG, LBP, and Gabor with a kNN classifier separately. The performance of the classifier was found on each feature set independent of the others with CK+ as well as the JAFFE database. The feature potential can be measured indirectly from the execution of the classifier as far as average recognition accuracy, overall accuracy, precision, recall and kappa coefficient. All these can be calculated from the confusion matrix, which reflects the number of correctly identified facial emotions along the diagonal. Sample confusion matrixes displaying the classifier performance with HOG and LBP features are displayed in Figure 6 for the CK+ database. Figure 6 (a) provides the kNN confusion matrix using the HOG feature vector, whereas Figure 6 (b) shows the confusion matrix using the LBP vector. Similarly, Figure 7 (a) displays the confusion matrix using Gabor features for the CK+ database whereas Figure 7 (b) shows the matrix for the JAFFE database. The confusion matrices have been computed for the kNN classifier for the JAFFE dataset with the HOG vector in Figure 8 (a) and LBP features in Figure 8 (b).



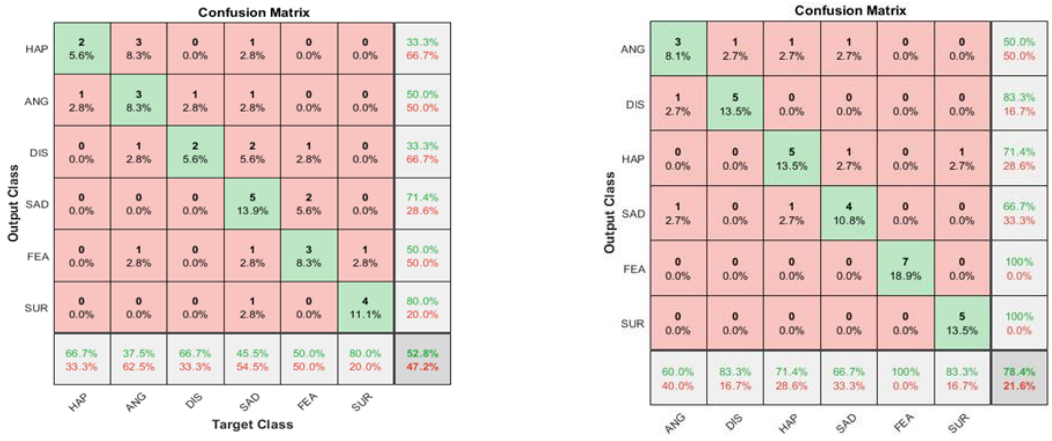
(a) (b)

Figure 6. The testing confusion matrix using kNN classifier for CK+ dataset using, (a) HOG feature set and (b) LBP feature set



(a) (b)

Figure 7. The testing confusion matrix using kNN classifier with Gabor feature set for, (a) CK+ dataset and (b) JAFFE dataset



(a) (b)

Figure 8. The testing confusion matrix using kNN classifier for JAFFE dataset using, (a) HOG feature set and (b) LBP feature set

The proposed schemes have been implemented on an intel ® core ™ i3-2330M CPU, 220 GHz laptop with 4GB RAM, and 64-bit OS using MATLAB R2018b. The various performance parameters used in this paper are defined. From this, we will better discriminate the features.

a. Overall Accuracy – it is the ratio of the number of correctly classified individuals to the total number of individuals tested.

$$\text{OA}= \frac{TruePositive + TrueNegative}{TruePositive+ TrueNegative+ FalsePositive+FalseNegative}$$

b. Average Accuracy – Average accuracy can be written as the sum of accuracies of each class divided by the total number of the available classes present.

c. Precision – Precision is given as,

$$\text{Precision} = \frac{TruePositive}{TruePositive+ FalsePositive}$$

d. Recall – it is given as,

$$\text{Recall} = \frac{TruePositive}{TruePositive+ FalseNegative}$$

e. Kappa Coefficient (K) – Kappa coefficient is a statistic used to measure the agreement between two or more observers [24]. The value of $K < 0$ conveys no unity, the gain lying 0–0.20 indicates low unity, the gain value lying 0.21–0.40 conveys good unity, the gain lying 0.41–0.60 gives moderate unity, the gain lying 0.61–0.80 gives substantial unity, and the gain lying 0.81–1 gives almost best unity [25].

f. Testing Time – Total time required for testing samples. Table 1 and Table 2 show the recognition accuracies of individual emotions for kNN based on three different feature schemes for CK+ and JAFFE datasets respectively. The surprise state has shown the highest accuracy using kNN+HOG and kNN+LBP as observed in Table 1. However, the fear and happy states have outperformed other chosen states using the Gabor+kNN for both the datasets when Table 1 and Table 2 are compared.

Table 3 and Table 4 display all the performance parameters of the three schemes used for the CK+ and JAFFE databases respectively. From these Tables, it is clear that the Gabor feature yields around 94% recognition accuracy which is the best among all and may be attributed to the image-enhancing capability of the Gabor transform, thus making the task easier for the classifier. The recognition accuracy of HOG based scheme is the lowest due to its limited structural information while the LBP-based scheme falls in between. It can also be observed that computation time has been highest for the scheme based on the Gabor feature because of its multi-resolution capability whereas it has been lowest for the HOG-based scheme.

Table 1. The percentage recognition accuracy of individual emotion for the CK+ database

Emotions	kNN+HOG	kNN+LBP	kNN+Gabor
Anger	87.5	92.5	92.5
Happy	83.9	94.4	100
Disgust	84.9	68.1	73.9
Sad	60.0	88.7	93.5
Fear	73.9	84.0	100
Surprise	94.7	98.6	96.0

Table 2. The percentage recognition accuracy of individual emotion for the JAFEE database

Emotion	kNN+HOG	kNN+LBP	kNN+Gabor
Anger	33.33	88.88	100
Sad	44.44	55.55	88.88
Happy	60.00	90.00	100
Disgust	12.50	87.50	87.50
Surprise	66.66	77.77	100
Fear	55.55	77.77	77.77

Table 3. Performance comparison of three different feature extractors for CK+ database

Feature	Overall accuracy	Average accuracy	Precision	Recall	Kappa Coefficient	Computation Time in sec.
HOG	84.53	80.81	0.80	0.82	0.80	2.1
LBP	90.65	90.31	0.90	0.92	0.88	2.2
Gabor	94.24	92.66	0.92	0.94	0.92	4.0

Table 4. Performance comparison of three different feature extractors for JAFFE database

Feature	Overall accuracy	Average accuracy	Precision	Recall	Kappa Coefficient	Computation Time in sec.
HOG	46.29	45.41	0.45	0.47	0.35	1.6
LBP	79.62	79.58	0.79	0.81	0.75	2.1
Gabor	92.59	92.36	0.92	0.94	0.91	4.5

5. CONCLUSION

This paper is an outcome of a survey conducted on three prominent feature extraction techniques used in PC vision and image processing issues for the task of emotion recognition from FE image datasets. The extracted feature sets from the JAFFE and CK+ datasets have been used to simulate the simple kNN classifier due to its ease of implementation and faster response. The application of the Gabor filter to binary images enhances the image to the desired standard, thus making the emotional models reliable and simple. Though there exist several challenges in the RFE system, a tremendous scope still exists. These developed models can be utilized effectively in automated teller machine (ATMs), identifying fake voters, passports, visas and driving licenses. It can also be applied in defense, identifying students in competitive exams as well as in private and government sectors. It can be inferred that the multi-resolution Gabor filters remain computationally expensive as compared to simple filters such as HOG and LBP, however, it has an improved recognition accuracy. The result can be extended in the future to other efficient feature extraction techniques that can describe facial expressive states adequately.

REFERENCES

[1] H. K. Palo and S. Sagar, "Characterization and Classification of Speech Emotion with Spectrograms," *Proceedings of the 8th International Advance Computing Conference, IACC 2018*, pp. 309–313, 2018, doi: 10.1109/IADCC.2018.8692126.

[2] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 1–12, 2015, doi: 10.1109/TAFFC.2014.2386334.

[3] D. Mehta, M. F. H. Siddiqui, and A. Y. Javaid, "Facial emotion recognition: A survey and real-world user experiences in mixed reality," *Sensors (Switzerland)*, vol. 18, no. 2, 2018, doi: 10.3390/s18020416.

[4] A. Alreshidi and M. Ullah, "Facial emotion recognition using hybrid features," *Informatics*, vol. 7, no. 1, 2020, doi: 10.3390/informatics7010006.

[5] S. Subudhiray, H. K. Palo, N. Das, and M. Chandra, "Comparison of Facial Emotion Recognition Using Effective Features," *Ann. Rom. Soc. Cell Biol.*, vol. 25, no. 6, pp. 5241–5252, 2021.

[6] A. Saxena, A. Khanna, and D. Gupta, "Emotion Recognition and Detection Methods: A Comprehensive Survey," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, pp. 53–79, 2020, doi: 10.33969/ais.2020.21005.

[7] N. Kumari and R. Bhatia, "Systematic review of various feature extraction techniques for facial emotion recognition system," *International Journal of Intelligent Engineering Informatics*, vol. 9, no. 1, p. 59, 2021, doi: 10.1504/ijiei.2021.116088.

[8] Z. Lei and G. Zhang, "Driver facial fatigue behavior recognition using local ensemble convolutional neural network and encoding vector," *Journal of Applied Science and Engineering*, vol. 23, no. 3, pp. 385–390, 2020, doi: 10.6180/jase.202009_23(3).0001.

[9] I. M. Revina and W. R. S. Emmanuel, "A Survey on Human Face Expression Recognition Techniques," *Journal of King Saud University - Computer and Information Sciences*, 2018, doi: 10.1016/j.jksuci.2018.09.002.

[10] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry*, vol. 11, no. 10, 2019, doi: 10.3390/sym11101189.

[11] N. S. Kumar and P. Praveena, "Evolution of hybrid distance based kNN classification," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 510–518, 2021, doi: 10.11591/IJAI.V10.I2.PP510-518.

[12] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors (Switzerland)*, vol. 18, no. 2, 2018, doi: 10.3390/s18020401.

[13] S. Subudhiray, H. K. Palo, N. Das, and S. N. Mohanty, "Comparative analysis of histograms of oriented gradients and local binary pattern coefficients for facial emotion recognition," *Proceedings of the 2021 8th International Conference on Computing for Sustainable Global Development, INDIACom 2021*, pp. 18–22, 2021, doi: 10.1109/INDIACom51348.2021.00005.

[14] S. Saha *et al.*, "Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm," *Applied Sciences (Switzerland)*, vol. 10, no. 8, 2020, doi: 10.3390/APP10082816.

[15] M. R. Mahmood, M. B. Abdulrazzaq, S. R. M. Zeebaree, A. K. Ibrahim, R. R. Zebari, and H. I. Dino, "Classification techniques' performance evaluation for facial expression recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2, pp. 1176–1184, 2020, doi: 10.11591/ijeecs.v21.i2.pp1176-1184.

[16] X. Zhao and S. Zhang, "A review on facial expression recognition: Feature extraction and classification," *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, vol. 33, no. 5, pp. 505–517, 2016, doi: 10.1080/02564602.2015.1117403.

[17] Z. Xiang, H. Tan, and W. Ye, "The Excellent Properties of a Dense Grid-Based HOG Feature on Face Recognition Compared to Gabor and LBP," *IEEE Access*, vol. 6, pp. 29306–29318, 2018, doi: 10.1109/ACCESS.2018.2813395.

[18] H. K. Meena, S. D. Joshi, and K. K. Sharma, "Facial Expression Recognition Using Graph Signal Processing on HOG," *IETE Journal of Research*, vol. 67, no. 5, pp. 667–673, 2021, doi: 10.1080/03772063.2019.1565952.

[19] D. Lakshmi and R. Ponnusamy, "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders," *Microprocessors and Microsystems*, vol. 82, 2021, doi: 10.1016/j.micpro.2021.103834.

[20] Y. Tong and R. Chen, "Local Dominant Directional Symmetrical Coding Patterns for Facial Expression Recognition," *Computational Intelligence and Neuroscience*, vol. 2019, 2019, doi: 10.1155/2019/3587036.

[21] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 7, pp. 1056–1064, 2008, doi: 10.1016/j.engappai.2007.11.010.

[22] L. L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and General Discriminant Analysis for face identification and verification," *Image and Vision Computing*, vol. 25, no. 5, pp. 553–563, 2007, doi: 10.1016/j.imavis.2006.05.002.

[23] K. S. Yadav and J. Singha, "Facial expression recognition using modified viola-john's algorithm and kNN classifier," *Multimedia Tools and Applications*, vol. 79, no. 19–20, pp. 13089–13107, May 2020, doi: 10.1007/s11042-019-08443-x.

[24] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The kappa statistic," *Family Medicine*, vol. 37, no. 5, pp. 360–363, 2005.

[25] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012, doi: 10.11613/bm.2012.031.

BIOGRAPHIES OF AUTHORS

Swapna Subudhiray received the M. Tech degree in Electronics and Communication Engineering in Electronics Communication Engineering from the Lovely Professional University, Jalandhar, India in 2012. She is currently working towards a Ph.D. degree at the Siksha 'O' Anusandhan (Deemed to be University), Odisha, India. Her current research interests include Image Processing, Machine Learning, Deep Learning, and Artificial Intelligence. She can be contacted at email: swapnaray89@gmail.com.

Hemanta Kumar Palo received a Master of Engineering from Birla Institute of Technology, Mesra, Ranchi in 2011 and a Ph.D. in 2018 from the Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. Currently, he is serving as an Associate Professor in the Department of Electronics and Communication Engineering at the Institute of Technical Education and Research, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India. His area of research includes signal processing, speech and emotion recognition, machine learning, and analysis of power quality disturbances. He can be contacted at email: hemantapalo@soa.ac.in.

Niva Das completed her B. Tech, M. Tech, and Ph.D. from N.I.T Rourkela and BPUT, India. She has been working as a professor in the Department of Electronics & Communication Engineering at Siksha 'o' Anusandhan University, Bhubaneswar, India since 2010. Her research interest includes pattern recognition, machine learning, blind signal processing, biomedical signal processing, and document processing. She can be contacted at email: nivadas@soa.ac.in.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 66~78
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp66-78

Neural network-based parking system object detection and predictive modeling

Ziad El Khatib[1], Adel Ben Mnaouer[1], Sherif Moussa[1], Omar Mashaal[1], Nor Azman Ismail[2], Mohd Azman Bin Abas[2], Fuad Abdulgaleel[2]
[1]Department of Electrical and Computer Engineering, Canadian University Dubai, Dubai, United Arab Emirates
[2]Department of Computer Science, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Article Info

Article history:

Received Nov 6, 2021
Revised Aug 31, 2022
Accepted Sep 19, 2022

Keywords:

Hyperparamter optimization
Predictive modeling
Real-time object detection
Regularization technique
Yolo neural network

ABSTRACT

A neural network-based parking system with real-time license plate detection and vacant space detection using hyper parameter optimization is presented. When number of epochs increased from 30, 50 to 80 and learning rate tuned to 0.001, the validation loss improved to 0.017 and training object loss improved to 0.040. The model means average precision mAP_0.5 is improved to 0.988 and the precision is improved to 99%. The proposed neural network-based parking system also uses a regularization technique for effective predictive modeling. The proposed modified lasso ridge elastic (LRE) regularization technique provides a 5.21 root mean square error (RMSE) and an R-square of 0.71 with a 4.22 mean absolute error (MAE) indicative of higher accuracy performance compared to other regularization regression models. The advantage of the proposed modified LRE is that it enables effective regularization via modified penalty with the feature selection characteristics of both lasso and ridge.

Corresponding Author:

Ziad El Khatib
Department of Electrical and Computer Engineering, Canadian University Dubai
City Walk, Dubai, United Arab Emirates
Email: Ziad.Elkhatib@cud.ac.ae

1. INTRODUCTION

With the rapid increase of car users the need for parking system with predictive modeling that allows learning by analyzing past user information is becoming essential. Prediction models are required since machine learning algorithms can predict future parking demand and the behavior of the parking users. Moreover, with trained neural network-based parking system license plate detectors, the parking system license plate detection and classification will improve with you only look once (YOLO) neural-network algorithm. YOLO provides real-time license plate detection [1]–[5]. Du *et.al.* [1] published work emphasis the need to have a multi-plate processing and detection [1] where YOLO can do that with increased speed and accuracy [6]–[8]. Masood *et.al.* published [9] work presents license plate detection and recognition using convolution neural network (CNN) with only 93.4% accuracy. Silva and Jung [3] published work presents license plate detection and recognition using CNN without optimizing neural-network hyperparameters [3]. Hyperparameter optimization is performed on the proposed general-purpose graphic unit (GPU)-based neural-network real-time parking system object detection. Also, other published work [4], [10] do not include hyperparameter optimization in their neural network processing. Nyambal and Klein their automated parking space detection using CNN achieving a 95.5% accuracy without license plate detection [10]. Fukusaki *et.al.* [11] also presented their published work on parking space detection using CNN without license plate detection [11]. Acharya *et.al.* [12] published work descirbes parking system with parking space neural network detection

achieving high accuracy of 99.7% without any predicitive modeling processing [12]. Lin *et.al.* [13] and Idris *et.al.* [14] and Sarangi *et.al.* [15] in their published work presented a survey of smart parking system without any proposed design implementation. Applying existing machine learning in smart parking applications is investigated in [16]–[20]. However, they look at the data analytics without proposing a system implementation. Other published work [21]–[23] do not propose a machine learning model algorithm. Simhon *et.al.* [24] present smart parking system with predictive modeling in their published work without neural network object detection [24]–[26] published work describes parking system with predicitive modeling without license plate and parking space neural network detection. Other published work [27], [28] propose parking system implementation without looking at the machine learning data analytics algorithms and do not propose predictive modeling in their post processing system.

2. REAL-TIME NEURAL NETWORK OBJECT DETECTION WITH HYPERPARAMETER OPTIMIZATION

Hyperparameter optimization is applied to determine the optimal values of hyperparmaters such as optimal learning rate and the number of epoch in order to improve precision and accuracy [29]–[32]. Figure 1 shows neural network-based parking system real-time license plate detection with YOLO. YOLOv5 which is based on PyTorch framework provides real-time object detection with high accuracy and speed [6], [7].

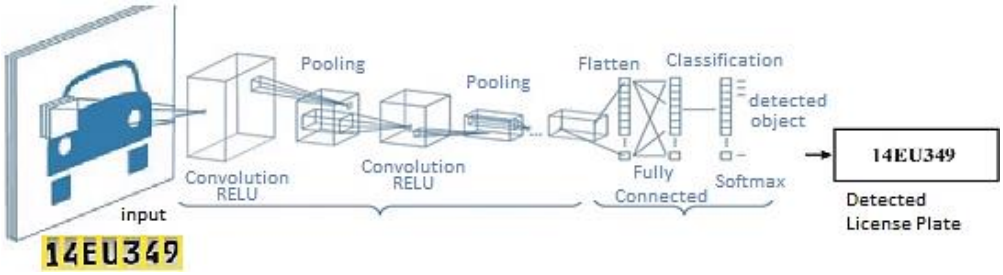


Figure 1. Neural network-based parking system real-time license plate detection

As shown in Figure 1 neural network-based parking system real-time license plate detection with YOLO is based on PyTorch framework provides real-time object detection with higher accuracy and speed [5], [7], [29]–[32]. YOLO takes the in a single instance by the framework and divides it into a grid with each grid having a dimension of n by n. Then places bounding box in the residual blocks and then determines the intersection over union (IOU). YOLO uses IOU to provide an output box that surrounds the object. YOLO then predicts the class probabilities for these boxes and their coordinates unlike CNN. After classification and localization are applied on each grid then the data that is labelled are passed to the model in order to train it. We determine YOLO loss function with (1).

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{\mathrm{B}} 1_{ij}{}^{obj} (x_i - x'_i)^2 + (y_i - y'_i)^2 \quad (1)$$

Then the bounding box location (x, y) is determined with (2), when there is object the $1_{ij}{}^{obj}$ is 1 and 0 when there is no object.

$$+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{\mathrm{B}} 1_{ij}{}^{obj} [\left(\sqrt{w_i} - \sqrt{w'_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{h'_i}\right)^2] \quad (2)$$

The bounding box size (w, h) when there is object can be determined with (3).

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{\mathrm{B}} 1_{ij}{}^{obj} (C_i - C'_i)^2 \quad (3)$$

The confidence when there is object is determined with (4).

$$+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{\mathrm{B}} 1_{ij}{}^{noobj} (C_i - C'_i)^2 \quad (4)$$

1_{ij}^{noobj} is 1 when there is no object, 0 when there is object. The class probabilities when there is object is determined with (5).

$$+\sum_{i=0}^{S^2} 1_{ij}{}^{obj} \sum_{c \in classes} \left(p_i(c) - p'_i(c)\right)^2 \quad (5)$$

3. REAL-TIME OBJECT DETECTION WITH HYPERPARAMETER OPTIMIZATION

In training neural network algorithm during learning model, it is important to look at the loss function in order to get intuition about how the neural network detection and classification are learning. Hyperparameter optimization is applied to determine the optimal values of hyperparmaters such as optimal learning rate and the number of epochs in order to improve precision and accuracy [33]. In training algorithm, epoch training setting start with 10 epochs then 50 and then 80 epochs. Epoch can be defined as how many times you pass once for learning the entire complete dataset through the neural network. The model is incrementally trained with more epoch which is increased in intervals of 10, 30, 50 and 80. Figure 2 shows the loss function with optimized hyperparameter learning rate as the number of epochs is increased. If we train the model with a lot of epochs this leads to overfitting of training model, whereas if we train the model with little epochs this leads to underfit model.

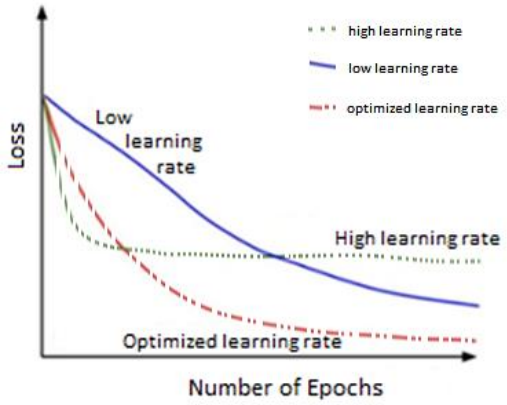


Figure 2. Loss function with optimized hyperparameter learning rate and number of epochs [33]

The learning-rate parameter decided how big step should be taken when searching for an optimal solution. The learning rate is tuned during hyperparameter optimization to improve the loss-function as shown in Figure 2 less learning-rate would require lots of epochs which increase the training time, however more learning-rate require fewer epochs. We adjust the learning rate during training incrementely from high to low once we get closer to the optimal solution. We adjust the learning rate during training from high to low once we get closer to the optimal solution.

Validation loss is the loss calculated on the validation set, when the data is split using cross-validation [33]. If validation loss gets worse that indicates overfitting as can be seen in Figure 2. As long as the validation loss and training loss continues to improve, we keep optimizing the hyperparameters. The objective is to make the validation loss as low as possible to improve model accuracy. Learning rate adjust the weights and it will converge slower with lower value of the learning.

One of the most common evaluation metrics that is used in neural network object detection is 'mAP', which stands for 'mean average precision. A good mAP indicates a stable consistent model. Figure 3(a) and Figure 3(b) show the precision and the mean average precision for both training loss and validation loss performance as epoch is increased. We want to make the validation loss as low as possible to improve model accuracy. Figure 3(a) shows precision performance and Figure 3(b) shows mAP performance as the number of epochs is increased. As can be seen in Figure 3 when number of epochs increased from 30 to 50 and then to 80 the model mean average precision mAP_0.5 is improved to 0.99 and the precision is improved to 99%. Figure 4(a) shows training object loss and Figure 4(b) shows training class loss performance as the number of epochs is increased.

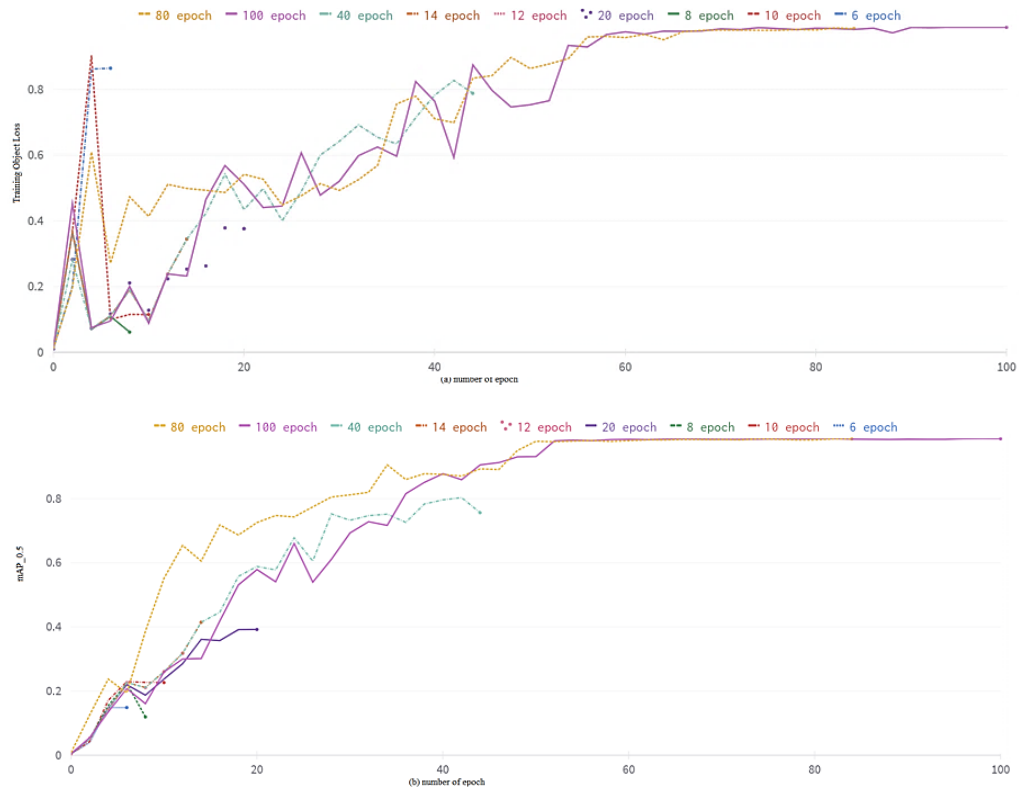


Figure 3. Accuracy performance as the number of epochs is increased (a) precision and (b) mAP

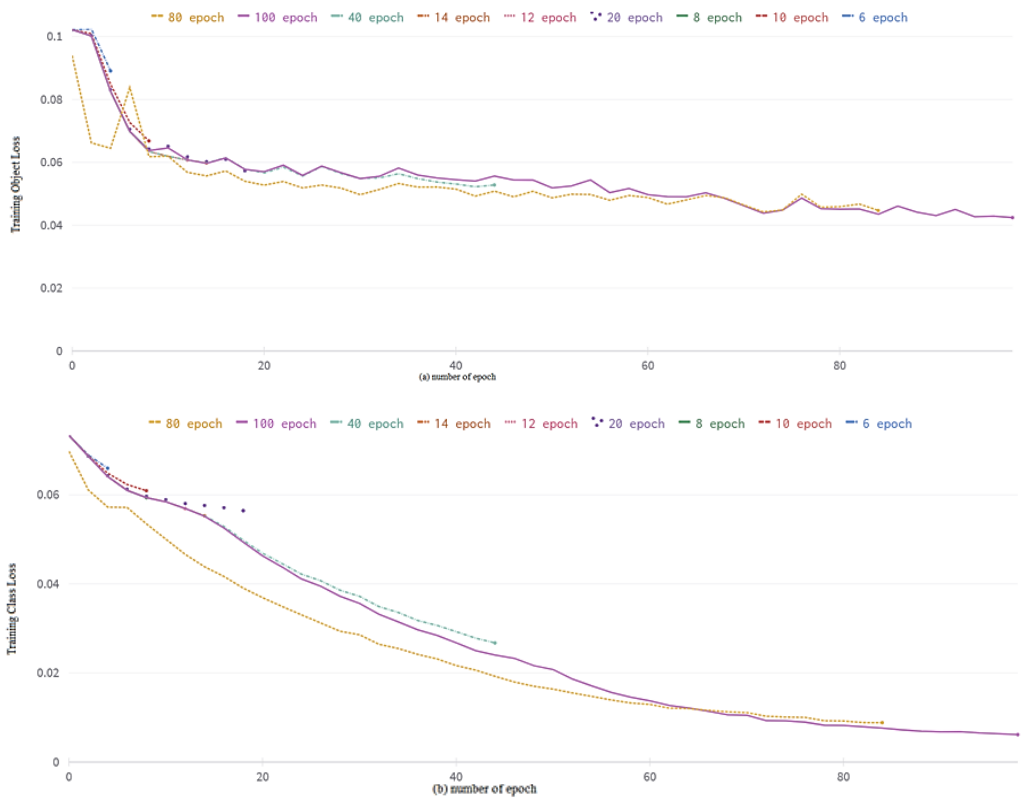


Figure 4. Accuracy performance as the number of epochs is increased (a) training object and (b) training class loss

As can be seen in Figures 5(a) and 5(b) when number of epochs increased from 30 to 50 and then to 80 and learning rate tuned to 0.001, the validation object loss improved to 0.017 and training object loss improved to 0.040. Figure 5(a) shows validation object loss and Figure 5(b) shows validation class loss performance as the number of epochs is increased. As can be seen in Figures 6(a) and 6(b) when number of epochs increased from 30 to 50 and then to 80 and learning rate tuned to 0.001, the validation box loss improved to 0.018 and training box loss improved to 0.017. Figure 6(a) shows training object loss and Figure 6(b) shows validation object loss performance as the number of epochs increased.

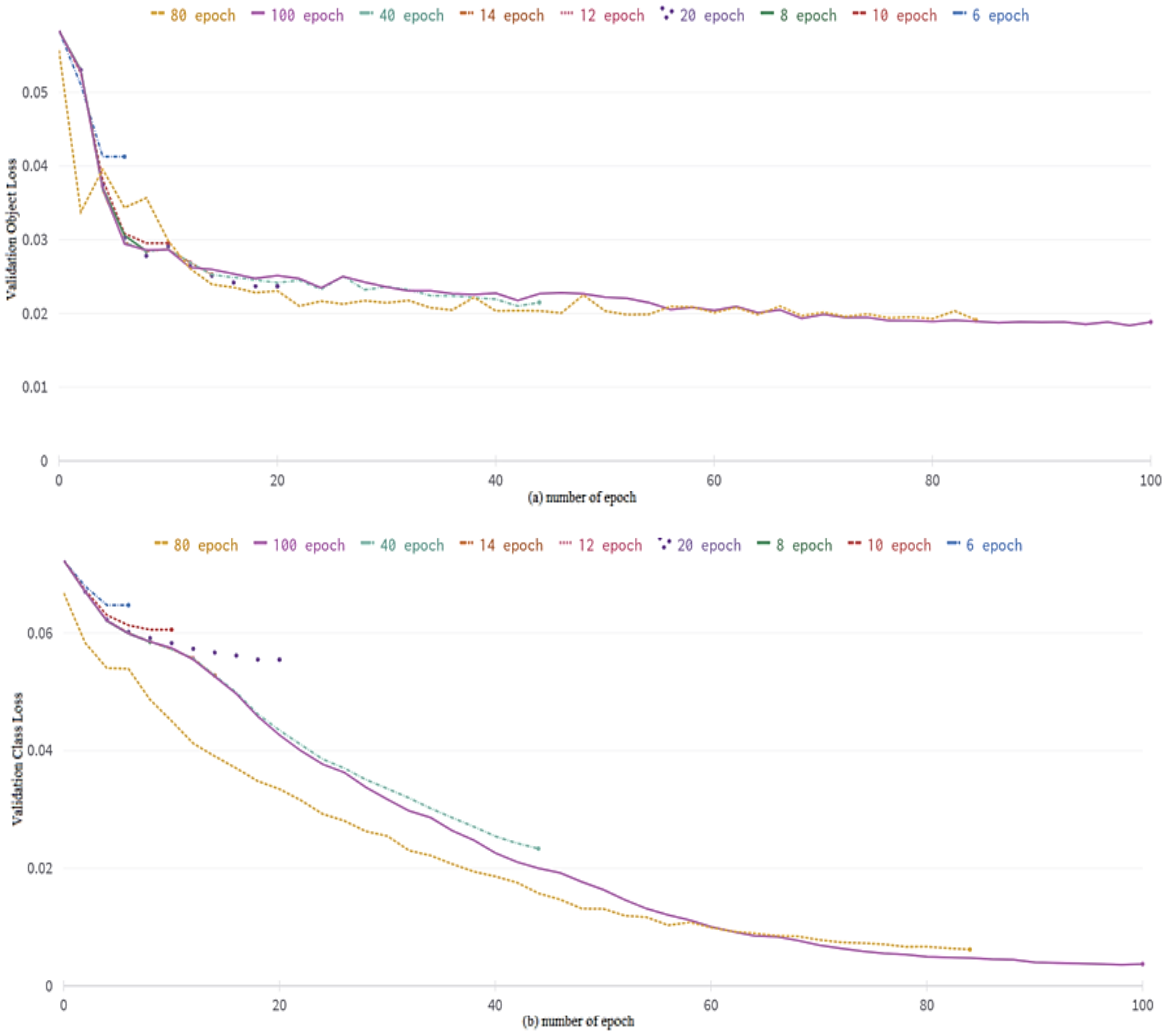


Figure 5. Accuracy performance as the number of epochs is increased (a) validation object loss and (b) validation class loss

The GPU-based neural network parking system real-time license plate detection with hyper parameter optimization model accuracy performance is shown in Table 1. When number of epochs increased from 30 to 50 and then to 80 and learning rate tuned to 0.001, the validation loss improved to 0.017 and training object loss improved to 0.040. Model mean average precision mAP_0.5 is improved to 0.988 and the precision is improved to 99%.

Table 1. Real-time neural-network object detection accuracy performance

Epoch	Learning Rate	mAP	Validation Object Loss	Training Object Loss	Precision
30	0.00001	0.652	0.030	0.060	0.50
50	0.0001	0.966	0.014	0.034	0.91
80	0.001	0.988	0.017	0.040	0.99

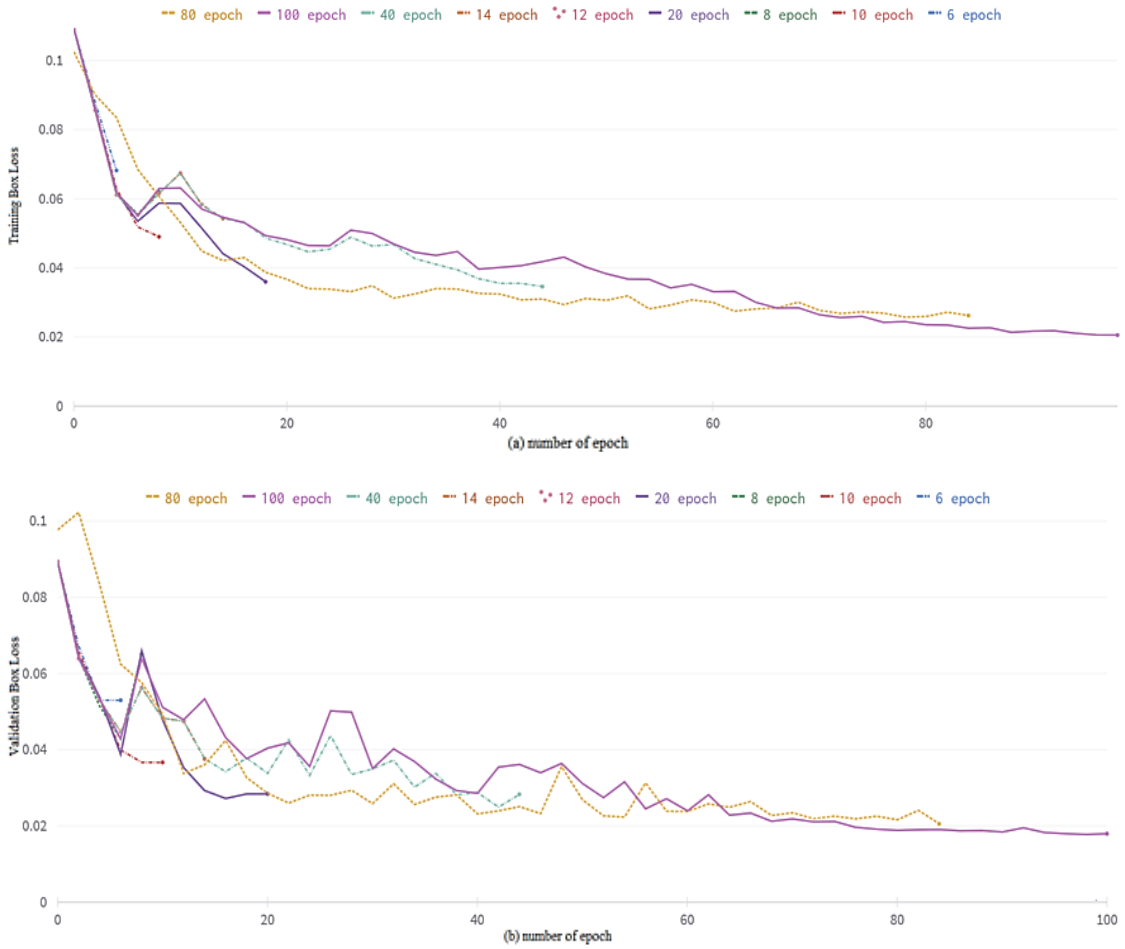


Figure 6. Accuracy performance as the number of epochs is increased (a) training box loss and (b) validation box loss

4. PREDICTIVE MODELING WITH REGULARIZATION TECHNIQUES

Figure 7 shows the proposed parking system with GPU processing using Nvidia Jetson Nano connected to artificial intelligence (AI) camera. The machine learning linear models provide a simple approach to predictive modeling. An overfit model is a model that fits the training dataset well but not the testing dataset as shown in Figure 8. The proposed parking system shown in Figure 7 uses Nvidia Jetson Nano connected to AI camera ce IMX 219 module. The AI camera connected to the Jetson board. A live video from the AI camera provides the real-time feed for vacant space detection. The 128-core Maxwell architecture-based GPU process the real-time processing analytics on the Jetson nano. Machine learning linear models provide a simple approach to predictive modeling [34]. An overfit model is a model that fits the training dataset well but not the testing dataset as shown in Figure 8. Overfitting causes low model accuracy. Regularization techniques solve the problem of overfitting [33], [35]–[37].

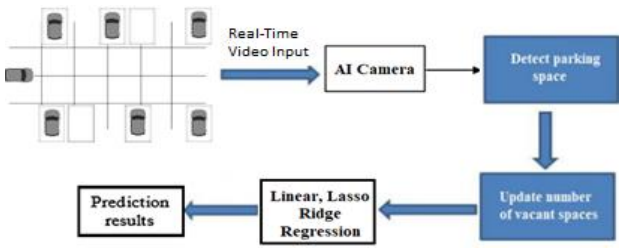


Figure 7. Proposed neural network-based parking system with real-time parking space detection and predictive modeling

4.1. Regularization techniques and overfitting

The objective is to have a machine learning model that has low bias and has low variability to produce consistent predictions across different datasets. Regularization is used in the proposed model to find the sweet spot between a simple model and a more complex model. Figure 8 shows the trade off between variance and bias in minimizing the prediction error [33], [35]–[37].

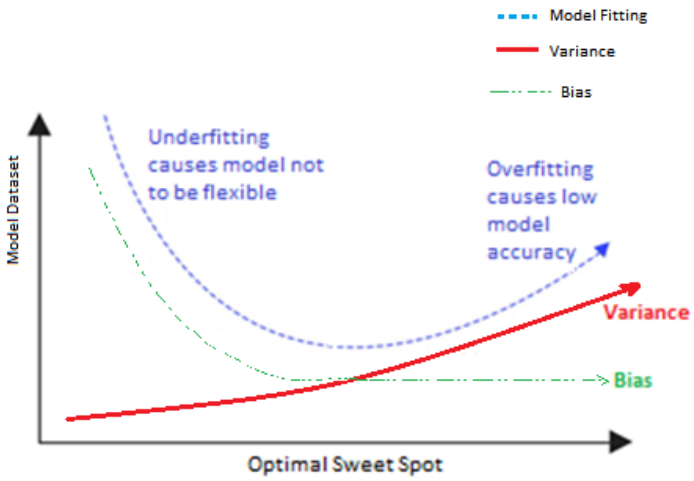


Figure 8. Regularization techniques solve the problem of overfitting

5. PROPOSED REGRESSION REGULARIZATION TECHNIQUE

Multiple linear regression uses a set of predictor variables and a response variable to fit a model of the form [33], [35]–[37].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \quad (6)$$

Y is the response variable and X is the predictor variable and βj is the slope. The values for β coefficients are chosen using the least square method which minimizes the residual sum of squares (RSS) [33], [35]–[37]. Regularization techniques function by penalizing the magnitude of coefficients along with minimizing the error between predicted and actual observations [38]. LASSO refers to least absolute shrinkable and selection operator [25]–[29], [31]. LASSO regularization is the process of adding a small modification to the cost function prevent the over-fitting problem as shown in (7) [33], [35]–[37].

$$J(m) = \Sigma\,(\text{yi} - \hat{\text{y}}\text{i})^2 + \lambda.|\text{slope}| \quad (7)$$

Where λ is the tuning parameter.
Least squares regression attempts to find coefficient estimates that minimize the RSS. The yi is the actual and ŷi is the predicted value for the ith observation based on the multiple linear regression model [33], [35]–[37].

$$RSS = \Sigma\,(\text{yi} - \hat{\text{y}}\text{i})^2 \quad (8)$$

Linear regression loss function is represented with Mean Squared Error function given by (9)

$$RSS = \frac{1}{n}\Sigma\,(\text{yi} - \hat{\text{y}}\text{i})^2 \quad (9)$$

Lasso is analogous to linear regression however it shrinks the coefficients of determination towards zero [33], [35]–[37]. Lasso lets you shrink and regularize these coefficients work on multiple datasets. Lasso regression seeks to minimize the following. Lasso lets you regularize these coefficients to work on different datasets. The second term in (5) is known as a shrinkage penalty. Lasso regression performs L1 regularization value. Ridge regularization is a variation of LASSO as the term added to the cost function is depicted (10) and (11). Ridge regression cost function model is given by[33], [35]–[37].

$$RSS + \lambda . \Sigma |\beta_j| \quad (10)$$

$$J(m) = \Sigma (yi - \hat{y}i)^2 + \lambda . |slope|^2 \quad (11)$$

Ridge regression instead tries to minimize (12)

$$RSS + \lambda . \Sigma \beta_j^2 \quad (12)$$

Ridge regression performs L2 regularization as shown in (7). A generalization of the ridge and lasso penalties, called the elastic net, combines the two penalties in (5) and (7). Elastic net regression seeks to minimize the following. The proposed modified lasso ridge elastic (LRE) regression model combines together both L1 and L2 regularization instead tries to minimize (13) and (14).

$$RSS + \lambda . \Sigma \beta_j^2 + \lambda . \Sigma |\beta_j| \quad (13)$$

$$RSS + \lambda^2 . \Sigma \beta_j^{3/2} + \lambda . \Sigma |\beta_j| \quad (14)$$

The advantage of the modified LRE penalty is that it enables effective regularization via modified penalty with the feature selection characteristics of lasso and ridge penalty.

6. REGRESSION MODAL PARTIAL DERIVATIVES

Linear regression equations needed to calculate the partial derivatives with respect to parameters of the loss function [38]. The values of model parameters m and b are updated using (15) and (16) [33], [35]–[37]. The updated values will be the values with which each step reduces the difference between the true and predicted values.

$$\frac{\partial}{\partial m} = \frac{2}{N} \sum_{i=1}^{N} -x_i \left(y_i - (mx_i + b)\right) \quad (15)$$

$$\frac{\partial}{\partial b} = \frac{2}{N} \sum_{i=1}^{N} -\left(y_i - (mx_i + b)\right) \quad (16)$$

The values of model parameters m and b are updated using (17) to (20) [33], [35]–[37]. The updated values will be the values with which each step reduces the difference between the true and predicted values. Ridge regression equations needed to calculate the partial derivatives with respect to parameters of the loss function [33], [35]–[37].

$$m = m - L_r . \frac{\partial L}{\partial m} \quad (17)$$

$$b = b - L_r . \frac{\partial L}{\partial b} \quad (18)$$

$$\frac{\partial L}{\partial \beta_o} = -\sum_{i=1}^{N} 2\left(y_i - \beta_o - \sum_{j=1}^{p} \beta_j x_j\right) \quad (19)$$

$$\frac{\partial L}{\partial \beta_j} = -\sum_{i=1}^{N} 2\left(y_i - \beta_o - \sum_{j=1}^{p} \beta_j x_j\right) x_i + 2\lambda\beta_j \quad (20)$$

The proposed modified LRE regression model equations needed to calculate the partial derivatives with respect to parameters of the loss function. The advantage of the modified LRE penalty is that it enables effective regularization via modified penalty with the feature selection characteristics of lasso and ridge penalty as shown in Figure 9. Jupyter python was used for coding the machine learning regularization regression modified LRE model algorithm. Figure 9 shows the proposed modified LRE regression model with dataset and with different tuning parameter.

$$\frac{\partial L}{\partial \beta_o} = -\sum_{i=1}^{N} 2\left(y_i - \beta_o - \sum_{j=1}^{p} \beta_j x_j\right) \quad (21)$$

$$\frac{\partial L}{\partial \beta_j} = -\sum_{i=1}^{N} 2\left(y_i - \beta_o - \sum_{j=1}^{p} \beta_j x_j\right) x_i + 1.5\lambda^2 \beta_j^{1/2} + \lambda \quad (22)$$

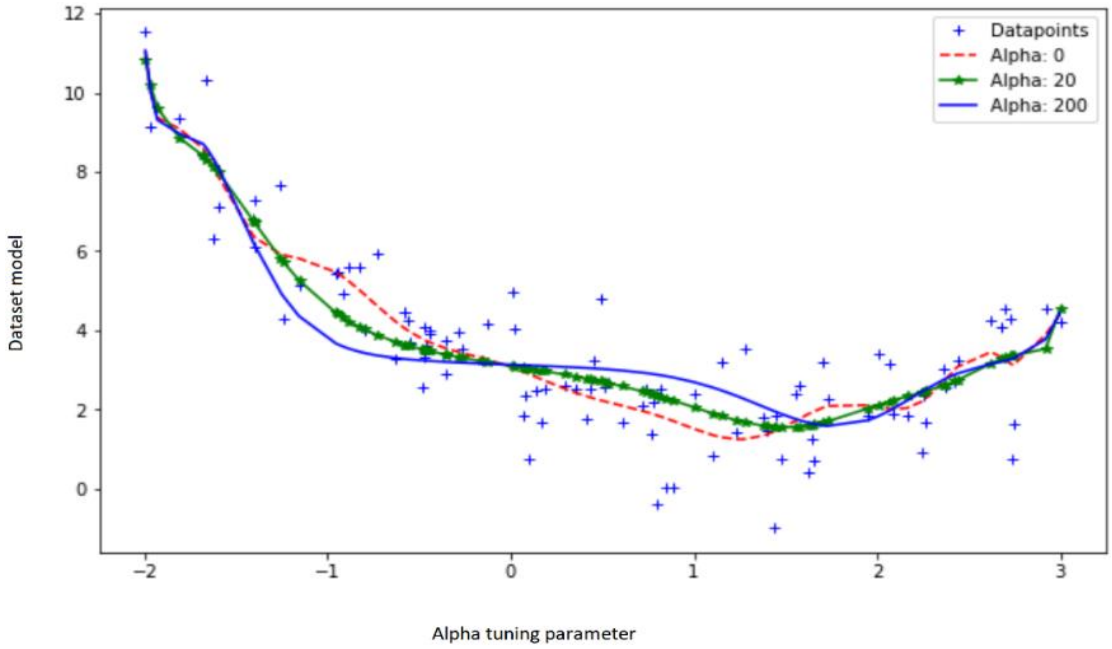


Figure 9. Proposed modified LRE regression model with dataset and with different tuning parameter

Figure 10 shows the linear regression predictive modeling forecasts in orange and green. The orange and green curves in both Figure 10 and Figure 11 are the future forcasts of the proposed predictive model the modified LRE indicating its effectiveness. Figure 11 shows proposed modified LRE regression predictive modeling forecasts in orange and green. Table 2 shows the proposed modified LRE regularization technique accuracy of 5.21 root mean square error (RMSE) and an R-Square of 0.71 with a 4.22 mean absolute error (MAE) compared to other regularization models.

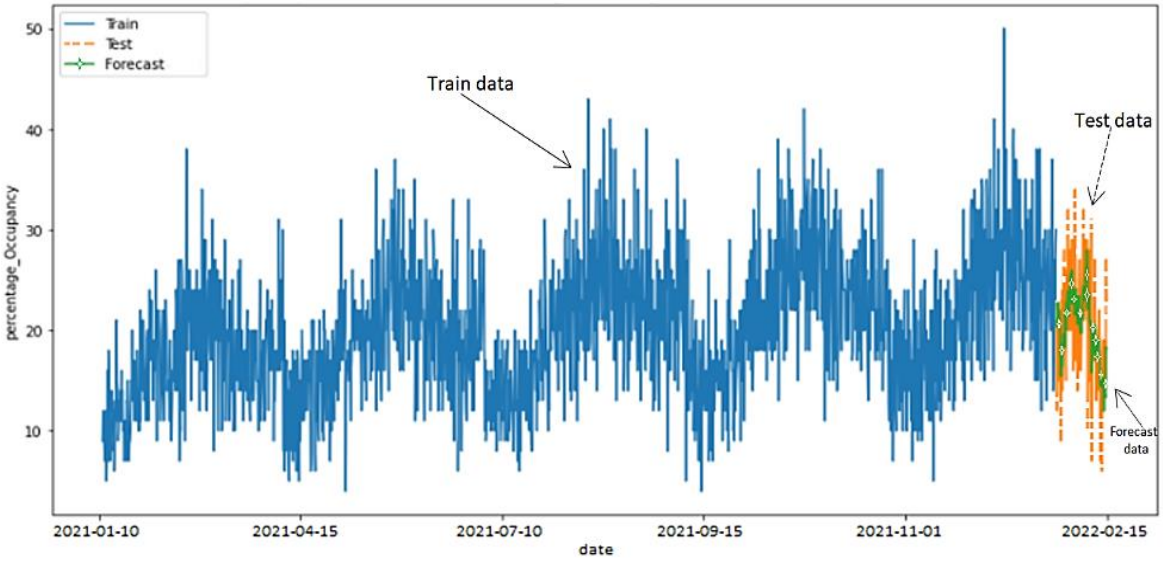


Figure 10. Linear regression predictive modeling forecasts in orange and green for test predictions

The neural network-based parking system with real-time license plate detection and vacant space detection using hyper parameter optimization is presented. When number of epochs increased from 30, 50 to 80 and learning rate tuned to 0.001, the validation loss improved to 0.017 and training object loss improved to 0.040. The model mean average precision mAP_0.5 is improved to 0.988 and the precision is improved to

99%. The proposed neural network-based parking system also uses a regularization technique for effective predictive modeling. The proposed modified LRE regularization technique provides a 5.21 RMSE and an R-square of 0.71 with a 4.22 MAE indicative of higher accuracy performance compared to other regularization regression models. The advantage of the proposed modified LRE is that it enables effective regularization via modified penalty with the feature selection characteristics of both lasso and ridge.

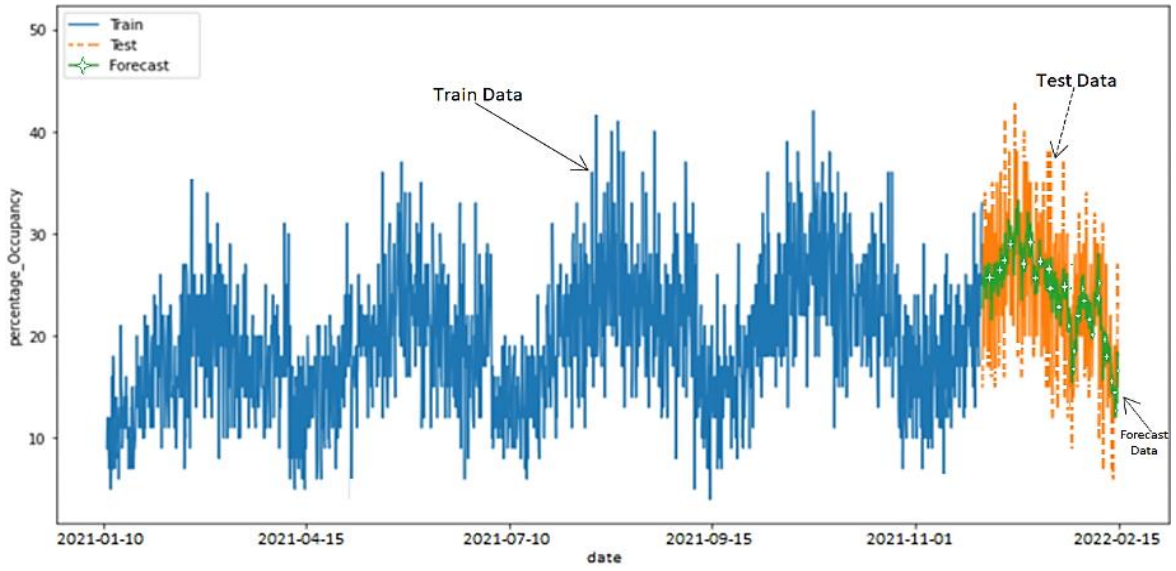


Figure 11. Proposed modified LRE regression predictive modeling forecasts in orange and green for test predictions

Table 2. Machine learning regression model accuracy performance for test predictions

Regression model	RMSE	R-Squared	MAE
Lasso Regression	5.45	0.66	4.52
Ridge Regresson	5.48	0.65	4.54
Elastic net Regresson	5.37	0.68	4.42
Linear Regression	5.72	0.63	4.67
Proposed Modified LRE Regression	5.21	0.71	4.22

7. CONCLUSION

The A neural network-based parking system with real-time license plate detection and vacant space detection using hyper parameter optimization has been presented. The model means average precision mAP_0.5 is 0.988 and the precision is 99%. The proposed neural network-based parking system uses a regularization technique for effective predictive modeling. The proposed modified LRE regularization technique provides a 5.21 RMSE and an R-square of 0.71 with a 4.22 MAE indicative of higher accuracy performance compared to other regularization regression models. The advantage of the proposed modified LRE is that it enables effective regularization via modified penalty with the feature selection characteristics of both lasso and ridge.

ACKNOWLEDGEMENTS

The authors thank all who contributed to the CUD-UTM research funding.

REFERENCES

[1] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 311–325, Feb. 2013, doi: 10.1109/TCSVT.2012.2203741.

[2] C. Gou, K. Wang, Y. Yao, and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted boltzmann machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1096–1107, Apr. 2016, doi: 10.1109/TITS.2015.2496545.

[3] S. M. Silva and C. R. Jung, "Real-time Brazilian license plate detection and recognition using deep convolutional neural networks,"

in *Proceedings - 30th Conference on Graphics, Patterns and Images, SIBGRAPI 2017*, Oct. 2017, pp. 55–62, doi: 10.1109/SIBGRAPI.2017.14.
[4] O. Bulan, V. Kozitsky, P. Ramesh, and M. Shreve, "Segmentation-and annotation-free license plate recognition with deep localization and failure identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2351–2363, Sep. 2017, doi: 10.1109/TITS.2016.2639020.
[5] R. Panahi and I. Gholampour, "Accurate detection and recognition of dirty vehicle plate numbers for high-speed applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 767–779, Apr. 2017, doi: 10.1109/TITS.2016.2586520.
[6] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, 2017, doi: 10.1109/CVPR.2017.690.
[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-Decem, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[8] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," *arXiv*, 2020, doi: 10.48550/arXiv.2004.10934.
[9] S. Z. Masood, G. Shu, A. Dehghan, and E. G. Ortiz, "License plate detection and recognition using deeply learned convolutional neural networks," *arXiv*, Mar. 2017, doi: 10.48550/arXiv.1703.07330.
[10] J. Nyambal and R. Klein, "Automated parking space detection using convolutional neural networks," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, Nov. 2017, vol. 2018-Janua, pp. 1–6, doi: 10.1109/RoboMech.2017.8261114.
[11] T. Fukusaki, H. Tsutsui, and T. Ohgane, "An evaluation of a CNN-based parking detection system with Webcams," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2020 - Proceedings*, 2020, pp. 100–103.
[12] D. Acharya, W. Yan, and K. Khoshelham, "Real-time image-based parking occupancy detection using deep learning indoor/outdoor seamless modelling, LBS, mobility view project The ISPRS benchmark on indoor modelling view project." 2018, [Online]. Available: https://www.youtube.com/watch?v=Ft94ypd4HxE.
[13] T. Lin, H. Rivano, and F. Le Mouel, "A Survey of smart parking solutions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3229–3253, Dec. 2017, doi: 10.1109/TITS.2017.2685143.
[14] M. Y. I. Idris, Y. Y. Leng, E. M. Tamil, N. M. Noor, and Z. Razak, "Car park system: a review of smart parking system and its technology," *Information Technology Journal*, vol. 8, no. 2, pp. 101–113, Feb. 2009, doi: 10.3923/itj.2009.101.113.
[15] M. Sarangi, S. K. Das, and K. S. Babu, "Smart parking system: survey on sensors, technologies and applications," in *1st IEEE International Conference on Advances in Information Technology, ICAIT 2019 - Proceedings*, Jul. 2019, pp. 250–255, doi: 10.1109/ICAIT47043.2019.8987378.
[16] N. Singh, S. Bawa, and H. Kaur, "Enhanced economy based smart parking system using machine learning," *Proceedings of Industry Interactive Innovations in Science, Engineering & Technology (I3SET2K19)*, 2020, doi: 10.2139/ssrn.3516600.
[17] J. Fan, Q. Hu, and Z. Tang, "Predicting vacant parking space availability: an SVR method with fruit fly optimisation," *IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1414–1420, Oct. 2018, doi: 10.1049/iet-its.2018.5031.
[18] Z. Zhao and Y. Zhang, "A comparative study of parking occupancy prediction methods considering parking type and parking scale," *Journal of Advanced Transportation*, vol. 2020, pp. 1–12, Feb. 2020, doi: 10.1155/2020/5624586.
[19] F. M. Awan, Y. Saleem, R. Minerva, and N. Crespi, "A comparative analysis of machine/deep learning models for parking space availability prediction," *Sensors (Switzerland)*, vol. 20, no. 1, p. 322, Jan. 2020, doi: 10.3390/s20010322.
[20] I. Masmoudi, A. Wali, A. Jamoussi, and A. M. Alimi, "Vision based system for vacant parking lot detection: VPLD," in *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, 2014, vol. 2, pp. 526–533, doi: 10.5220/0004730605260533.
[21] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, "Car parking occupancy detection using smart camera networks and deep learning," in *Proceedings-IEEE Symposium on Computers and Communications*, Jun. 2016, vol. 2016-Augus, pp. 1212–1217, doi: 10.1109/ISCC.2016.7543901.
[22] Q. Wu, C. Huang, S. Y. Wang, W. C. Chiu, and T. Chen, "Robust parking space detection considering inter-space correlation," in *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, ICME 2007*, Jul. 2007, pp. 659–662, doi: 10.1109/icme.2007.4284736.
[23] S. Lee, D. Yoon, and A. Ghosh, "Intelligent parking lot application using wireless sensor networks," in *2008 International Symposium on Collaborative Technologies and Systems, CTS'08*, May 2008, pp. 48–57, doi: 10.1109/CTS.2008.4543911.
[24] E. Simhon, C. Liao, and D. Starobinski, "Smart parking pricing: a machine learning approach," in *2017 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2017*, May 2017, pp. 641–646, doi: 10.1109/INFCOMW.2017.8116452.
[25] J. C. Provoost, A. Kamilaris, L. J. J. Wismans, S. J. van der Drift, and M. van Keulen, "Predicting parking occupancy via machine learning in the web of things," *Internet of Things (Netherlands)*, vol. 12, p. 100301, Dec. 2020, doi: 10.1016/j.iot.2020.100301.
[26] J. Barker and S. U. Rehman, "Investigating the use of machine learning for smart parking applications," Oct. 2019, doi: 10.1109/KSE.2019.8919291.
[27] A. Zacepins, V. Komasilovs, and A. Kviesis, "Implementation of smart parking solution by image analysis," in *VEHITS 2018 - Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, 2018, vol. 2018-March, pp. 666–669, doi: 10.5220/0006629706660669.
[28] Y. Ma, Y. Liu, L. Zhang, Y. Cao, S. Guo, and H. Li, "Research review on parking space detection method," *Symmetry*, vol. 13, no. 1, pp. 1–18, Jan. 2021, doi: 10.3390/sym13010128.
[29] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2017, vol. 2017-Octob, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.
[30] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object detection using convolutional neural networks," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, Oct. 2019, vol. 2018-Octob, pp. 2023–2027, doi: 10.1109/TENCON.2018.8650517.
[31] Ulagamuthalvi, J. B. J. Felicita, and D. Abinaya, "An efficient object detection model using convolution neural networks," in *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, Apr. 2019, pp. 142–147, doi: 10.1109/ICOEI.2019.8862698.
[32] M. A. Shehab, A. Al-Gizi, and S. M. Swadi, "Efficient real-time object detection based on convolutional neural network," in *2021 International Conference on Applied and Theoretical Electricity, ICATE 2021 - Proceedings*, May 2021, pp. 1–5, doi: 10.1109/ICATE49685.2021.9465015.

[33] P. Bühlmann and T. Hothorn, "Boosting algorithms: Regularization, prediction and model fitting," *Statistical Science*, vol. 22, no. 4, pp. 477–505, Nov. 2007, doi: 10.1214/07-STS242.
[34] F. J. W. M. Dankers, A. Traverso, L. Wee, and S. M. J. van Kuijk, "Prediction modeling methodology," in *Fundamentals of Clinical Data Science*, Springer International Publishing, 2018, pp. 101–120.
[35] D. Schreiber-Gregory and H. M. Jackson Foundation, "Regulation techniques for multicollinearity: lasso, ridge, and elastic nets," in *Proceedings of the SAS Conference Proceedings: Western Users of SAS Software*, 2018, pp. 1–23.
[36] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.
[37] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010, doi: 10.18637/jss.v033.i01.
[38] A. V Dorugade and D. N. Kashid, "Alternative method for choosing ridge parameter for regression," *Applied Mathematical Sciences*, vol. 4, no. 9, pp. 447–456, 2010.

BIOGRAPHIES OF AUTHORS

Dr. Ziad El-Khatib PhD in Electrical and Computer Engineering from Carleton University Canada. Assistant professor at Canadian University Dubai. Dr. Ziad El-khatib received his Bachelor of Science in Electrical Engineering from University of Ottawa Canada and his M.A.Sc. and PhD in Electrical and Computer Engineering from Carleton University Canada. He has several years of design experience in the field of communication integrated circuits at various companies including Nortel Networks Harris Corporation Chrysalis-ITS Semiconductor Itron Inc. and was adjunct professor in USA. He is currently assistant professor in the faculty of Electrical and Computer Engineering at Canadian University Dubai. His research interests include silicon based integrated circuits for radio frequency and microwave communications and power electronics integrated circuits. He has a book published through Springer on radio frequency amplification and linearization techniques and numerous IEEE journal and conference papers. He can be contacted at email: ziad.elkhatib@cud.ac.ae.

Dr. Adel Ben Mnaouer PhD in Computer Engineering Networking from Yokohama University Japan. Professor at Canadian University Dubai. Dr. Mnaouer holds a PhD in Computer Engineering Networking from Yokohama National University, Yokohama, Japan. He also obtained a Master of Engineering (Petri Nets) from Fukui University, Japan and a BSc in Computer Science from Ecole Superieur de Communications de Tunis. Prior to joining CUD, Dr. Mnaouer was Associate Professor and Vice Dean of Research at Dar Al Uloom University, Saudi Arabia. Prior to this he held academic posts at a range of institutions, including the University of Trinidad and Tobago, Carthage University, Tunis, Nanyang Technological University, Singapore and Sultan Qaboos University, Singapore. He can be contacted at email: adel@cud.ac.ae.

Dr. Sherif Moussa PhD in Electrical and Computer Engineering from University of Quebec Trois-Riviers, Canada. Associate professor at Canadian University Dubai. Dr. Sherif Moussa received his PhD in Electrical and Computer Engineering from University of Quebec Trois-Riviers, Canada, and his MSc degree in Electrical and Computer Engineering form University of Waterloo, Canada. His research areas are wireless communication, computer networks, and VLSI design. His research specifically focuses on MIMO-OFDM algorithms, multiple access OFDM, FPGA design and optimization. Dr. Moussa joined CUD in 2007 where he currently is working as an Assistant Professor at School of Engineering. Prior to joining CUD, he was a lecturer at School of Engineering, Centennial College, Toronto, Canada. Dr. Moussa is currently an active researcher who published in many international journals and conferences related to his field and he also currently serve as a reviewer and technical committee member for many international conferences. Dr. Moussa is the winner of 2015 CUD research excellence award and the founder of the flagship CUD robotics club. He can be contacted at email: smoussa@cud.ac.ae.

Omar Mashaal Masters in Communications Engineering from University of Technology Malaysia. Lecturer at Canadian University Dubai. Mr. Mashaal holds M.Eng in Communication Engineering from University of Technology, Malaysia and a BSc in Electrical Engineering - Communication Engineering from Ajman University. He worked as a voice and network engineer for two years and attained different industrial certificates from Alcatel-Lucent and Avaya. Mr. Mashaal attended several technical trainings, and he is a certified security intelligence analyst–educator by IBM. Mr. Mashaal is a member of the Jordan Engineers Association. His research interests are in antenna engineering and communication systems. He can be contacted at email: omar.mashaal@cud.ac.ae.

Dr. Nor Azman Ismail PhD in Computer Science and Computer Engineering from Loughborough. Associate Professor at University Teknologi Malaysia. Deputy Director of Office of Corporate Affairs (Web Director) and an academic staff at Computer Graphics and Multimedia Department, Universiti Teknologi Malaysia (UTM) for about thirteen years. Prior to my appointment as a University Web Director, He was Research Coordinator of Computer Graphics and Multimedia Department. In the earliest stages of my career, He was employed by Conner Peripherals a company that manufactured hard drives for personal computers and then Perak Department of Education. He can be contacted at email: azman@utm.my.

Dr. Mohd Azman bin Abas PhD in Computer Science and Computer Engineering from University Teknologi Malaysia. Professor at University Teknologi Malaysia. Director, Automotive Development Centre (ADC), Institute for Vehicle Systems and Engineering (IVeSE) Universiti Teknologi Malaysia (UTM). Research areas/interest, internal combustion engine, vehicle drive cycle, vehicle driving behaviour, autonomous driving behaviour, diver behaviour, motorsports. He can be contacted at email: azman.abas@utm.my.

Dr. Fuad Abdulgaleel PhD in Computer Science and Computer Engineering from University Teknologi Malaysia. Professor at University Teknologi Malaysia. Senior Lecturer at School of Computing, Faculty of Engineering and member in the Information Assurance and Security Research Group (IASRG). My research interests are – cyber threat intelligence, network security, misbehavior detection, anomaly detection, and malware analysis. He can be contacted at email: mdfarid@utm.my.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 79~86
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp79-86

Human emotion detection and classification using modified Viola-Jones and convolution neural network

Komala Karilingappa[1], Devappa Jayadevappa[2], Shivaprakash Ganganna[3]
[1]Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, India
[2]Department of Electronics and Instrumentation Engineering, Jagadguru Sri Shivarathreeshwara Academy of Technical Education, Bengaluru, India
[3]Department of Electronics and Instrumentation Engineering, M.S. Ramaiah Institute of Technology, Bengaluru, India

Article Info

Article history:

Received Oct 20, 2020
Revised May 26, 2022
Accepted Jun 24, 2022

Keywords:

Convolution neural network
Facial emotion recognition
Gray-level co-occurrence matrix
Linear binary pattern
Robust principal components analysis
Viola-Jones

ABSTRACT

Facial expression is a kind of nonverbal communication that conveys information about a person's emotional state. Human emotion detection and recognition remains a major task in computer vision (CV) and artificial intelligence (AI). To recognize and identify the many sorts of emotions, several algorithms are proposed in the literature. In this paper, the modified Viola-Jones method is introduced to provide a robust approach capable of detecting and identifying human feelings such as angerness,sadness, desire, surprise, anxiety, disgust, and neutrality in real-time. This technique captures real-time pictures and then extracts the characteristics of the facial image to identify emotions very accurately. In this method, many feature extraction techniques like gray-level co-occurrence matrix (GLCM), linear binary pattern (LBP) and robust principal components analysis (RPCA) are applied to identify the distinct mood states and they are categorized using a convolution neural network (CNN) classifier. The obtained outcome demonstrates that the proposed method outperforms in terms of determining the rate of emotion recognition as compared to the current human emotion recognition techniques.

Corresponding Author:

Komala Karilingappa
Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology
Tumakuru, Karnataka, India
Email: komalak@ssit.edu.in

1. INTRODUCTION

The human facial expression is one of the most essential and effective component of Inter-personnel communique. Facial expressions are quite costly. There is merely 7% of the total significant in a verbalized part of the message, 38% of the total signal in tone and 55% in portrayed [1]–[3]. Extracted features is very extensively utilized in surveillance, biometric, psychiatric, military and human-computer interaction (HCI) [4]. Facial images are exploited to recognize the type of emotion in humans. Anger, sadness, happiness, surprise, fear, disgust, and neutral are the seven primary emotions. Human facial expressions [5]–[8], can be used to identify the aforementioned states of emotion. Recognizing the human feelings is the important task. Several researchers have worked on the detection of age, sex and feelings from facial features [9]. Detection of different human emotions using facial expressions is a difficult task. The capacity of the system to differentiate between several faces is a frequent requirement in human-computer interaction. Until recently, computer vision issues were extremely difficult. With the advent of technology, the challenges in computer vision (CV) due to changes in lighting, ageing, hair, and other accessories [10], have become uncomplicated. Face recognition software, on the other hand, is used to enhance ease of access by identifying and verifying individuals based on their facial

attributes.Thus understanding facial attributes is vital for CV-based applications. These attributes and expressions helps for classifying the facial emotions. Artificial intelligence (AI) systems are employed on the basis of current technology innovations since these systems are capable of identification of emotion through facial characteristics [11]. Human emotion detection is still an active research area because of the current technology innovations for HCI in deep learning or convolution neural network (CNN) prototypes [12]–[14]. Various techniques are necessary to detect and categorize human faces, but deep learning methodology is better than other methods because of its huge capacities of varied datasets and quick computing capabilities [15]. Typically, the face recognition and classification involves several phases such as pre-processing, detection, feature extraction and classification. A voila-jones (V-J) technique is used for extracting the features by classifying images with emotion. This is usually followed by emotion classification using Haar and CNN [16]–[18]. The representation of extracted facial images with databases is the main shortcoming for the analyzying the features of lips and eye and the 2-D image. To overwhelme this shortcoming, the extracted images can be investigated with region of interest (ROI) [19]. Facial expression recognition (FER) can be done using statistical-unsupervised techniques like Independent component analysis (ICA) and genetic algorithm. Genetic algorithm is a feature enhancing technique that carry out for foreseeing Facial emotions [20]. Around 55% of total facial emotions is verified to contribute for social connections. Some of the limitations of the V-J algorithm includes a lack of accurate face and facial part recognition owing to lighting and variation issues. It also suffers from an inability to recognize a face and facial parts due to a fast shift in scene illumination and being too sensitive to stiff features in pictures. With low-resolution pictures and uneven lighting variations of the images, the updated algorithm V-J recognizes the face and facet part closely [21]. With an extremely low fictional rate and a high real-time video detection rate, it is quite resilient. It was suggested that the eye and mouth features are very important facial features which the algorithm extracts very effectively. When it comes to detecting different human emotions, it's quite accurate.

2. PROPOSED METHODOLOGY

In the proposed work, a distinctive technique is used for FER system using CNN. It consists of 3 important phases; face recognition, feature extraction followed by emotion classification. A video is taken as an input where the images can be extracted from the input video and then pre-process each of the images. The Gabor filter is used for removing the unwanted noise, blur and shadow from the original images. After pre-processing, the face detection is carried out using the modified V-J algorithm. There are four stages present in the modified V-J algorithm namely, Haar feature selection, Integral Image creation, AdaBoost training and cascading classifier. The Haar-feature is useful to apply on input face images to check whether the faces are present or not in an image. It can be computed as result of addition of all image pixels, and then subtracted to obtain a unique value. If the unique value is greater than the range, then it implies the human face is recognized. Creation of Integral Image is used for evaluating the sum-up of pixels in a particular area of interest of an image. Adaboost is used for generating the robust classifiers from feasible classifiers. It is not only used to reduce the detection of false positive rate but also decreases the difficulty due to the presence of redundant features. Cascade structure is not only utilized for removal of the false positive images as well as utilized to inspect the occurrence of a face in a specific part of an image. This is followed by extraction of features from the image by gray-level co-occurrence matrix (GLCM) and linear binary pattern (LBP). Afterwards, the required feature is selected using principle component analysis (PCA). The particular features are fed to the CNN classifier for classification. The output from the CNN classifier is the type of emotion in the image in question.

The most important phase in FER is face detection to identify all emotions efficiently using the modified V-J algorithm. The face and emotion can be detected using the proposed algorithm. Extracting the features plays a importance in the FER system as a result of enhancing the accuracy of Feeling detection techniques. There are many extraction techniques such as LBP, GLCM, gray level weight matrix (GLWM), traditional gabor filter (TGF) and daubechies wavelet packet features (DBWP). In the proposed methodology, extraction of feature techniques such as GLCM and LBP are used for classifying the texture. Using GLCM dissimilarity, correlation, mean, entropy, variance, average angular second moment, homogeneity, contrast, energy, standard deviation and maximum probability features are extracted. LBP is used as texture operator which symbols the image pixels through adopting the process of thresholding the neighborhood of each pixel. The output of LBP is obtained in the form of binary. Due to discernment of power and computational simplicity [22], LBP is a widely used method in real time applications. The popularity of LBP is due to its robustness toward monotonic varies in gray-scale due to light illumination change. In LBP every pixel value p is compared with the radial distance r of its N neighbors. There are N comparisons for each pixel p and the output for each can be expressed as:

$$LBP\ (m,n)\ =\ \sum_{p=0}^{7} s\ x(\text{gc-gp})2^{l} \quad (1)$$

where, ‘gc’ corresponds to the value of grey scale in the centrally located pixel (xc, yc) and ‘gp’ to the grey scale values of the eight

$$s(z) = \begin{cases} 1, z\ \geq\ 0 \\ 0, z\ <\ 0 \end{cases} \quad (2)$$

neighboring pixels. p is the number of neighboring pixels, s(z) is a threshold function. After feature extraction, feature selection is used to upgrade the performance of the classifier. robust principal components analysis (RPCA) technique is employed for extracting the features from face images and also used to reduce the dimensionality of face images. It is a numerical method that tranforms a set of correlated N face images to a set of eigen face images.

The RPCA was formulated as a non-convex optimization problem defined as,

$$arg\ \ \min_{L,S} rank(L) + \lambda||S||_0 \text{ s.t D=L+S} \quad (3)$$

A set of face images are in training, then it is denoted with large eigenvalues through the greatest eigenfaces for accurate estimation of the face. After this step the result of eigenfaces, each face image can be indicated by permutation of eigenfaces, followed by symbolization in the form of vectors. The input features are compared with standard features of dataset for FER. The features are classified using CNN classifier. CNN comprises sequences of convolutional layers, the output which is correlated only to native areas in the input. This is carried-out through sliding filter, or weighted-matrix with respect to the input. For every point, CNN computing the product of convolution between the input and filter [23], [24].

Figure 1 show the block diagram of the proposed FER system. Initially from real time video, facial image will be captured than fed to pre-processing. In the next stage face detection is done using modified V-J method. Facial feature extraction is done using GLCM and LBP. These methods were also used to distinguish the texture information of images and hence it improves the classification performance. The feature selection using RPCA method is done. The RPCA is a feature selection technique which is used for ease the dimensionality of face data. This step is followed by feeding the image to CNN classifier, where the real time image will be compared with database to detect the facial expression more effectively. Figure 2 shows the flowchart of the proposed methodology, which is self explanatory.

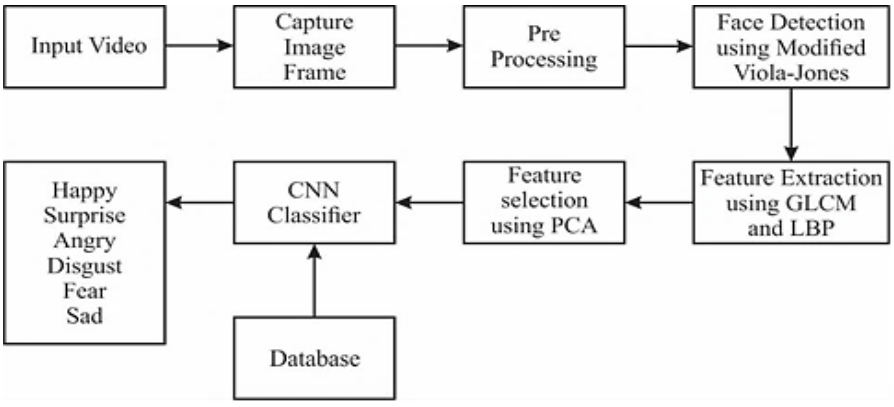


Figure 1. Block diagram of the proposed FER system

3. RESULTS AND DISCUSSION

The proposed work is implemented using the technical computing environment MATLAB. The datasets were collected from Kaggle and karolinska directed emotional faces (KDEF) databases [25]. This dataset comprises 215 images with 7 facial emotions such as happy, sad, surprise, disgust, angry, fear and neutral. The real-time images are used as input images. In the beginning, pre-processing step is used for removal of unwanted images as well it smoothen the images from input datasets using the Gabor filter. For FER, the modified V-J algorithm is used to vary the image intensity and window size. The AdaBoost is not only used to reduce the detection of false positive rate but also decreases the difficulty due to the presence of

redundant features. CNN classifiers are used to classify effectively the different emotion statuses of the input images effectively. The proposed technique yielded an accuracy validation of 95.6%.

The Kaggle and KDEF databases are used as shown in Figure 3, the training and testing sets can be divided through cross-validation. In this validation, the whole database is segregated into three identical sets of images. The segregation is random in nature. Then two sets are combined to use as a training data set. The remaining section of the dataset is used for the testing phase. Figure 4 shows the accuracy and log loss plot of CNN during Training. From figure one can notice the quality of performance of a model as the number of iterations of optimization progresses. Accuracy metric is used to measure the performance in an interpretable way. It is a degree of how accurate the model's likelihood is compared to the correct data. Figure 5 shows Adaboost plot, which gives the relationship between false positive rate and true positive rate. Adaboost is used for adjusting the weights of classifiers during training. The process is repeated as training process iterates. This step ensures that the accuracy of predictions of unusual observation. It is also used to boost the performance of any machine learning algorithm. Figure 6 shows the bargraph of performance of different classifiers for the selected dataset. We can see that CNN has higher performance value among the compared classifiers for the chosen comparision matrix.

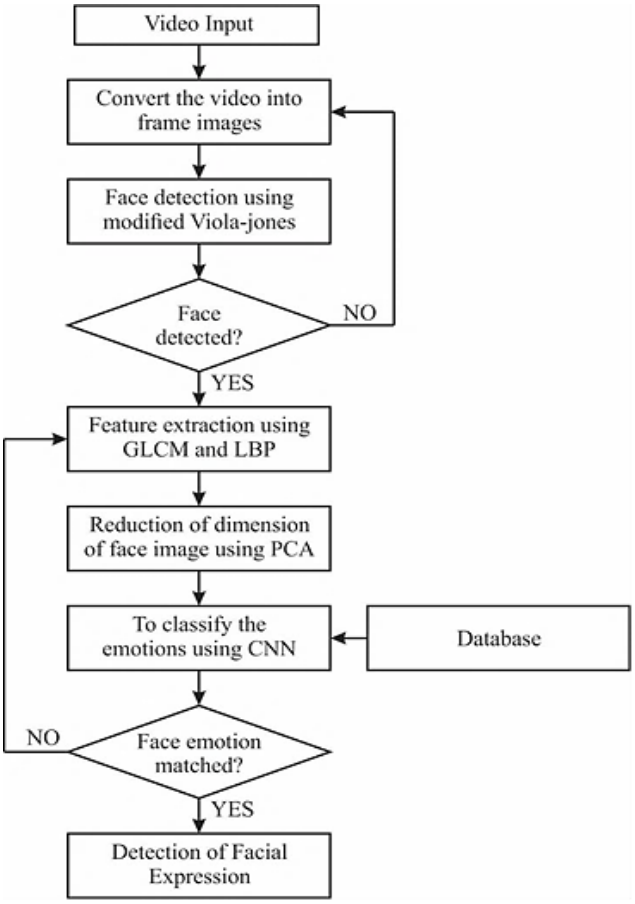


Figure 2. Flowchart of proposed methodology

3.1. Performance analysis

The performance of the proposed work is quantitatively evaluated using the parameters like precision, sensitivity, specificity, accuracy and recall. The confusion matrix of the facial emotion detection is constructed as shown in Table 1 for the merged image. The experimental results show that the proposed technique efficiently detects the facial expressions with high accuracy as compared to the current techniques. Table 2 shows the region of interest and its corresponding real time image. The classification of emotion is displayed above the input image. Table 3 shows the accuracy outcome of CNN classifiers to be more effective in detecting emotions compared to k-nearest neighbor (KNN) and artificial neural network (ANN) classifiers.

Figure 3. Sample dataset used for the proposed work

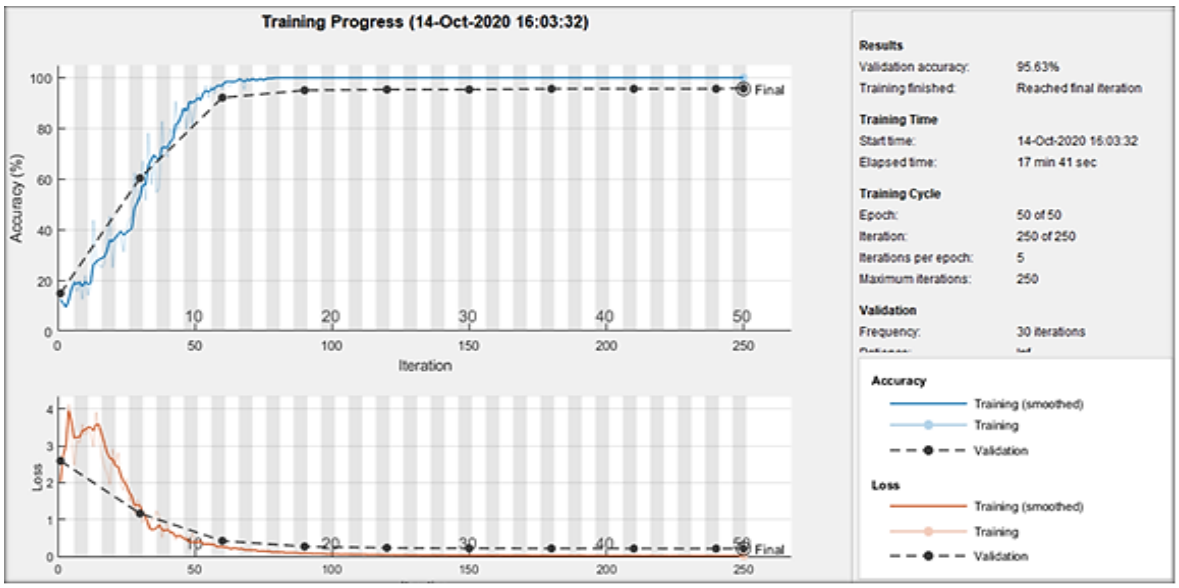


Figure 4. Accuracy and log-loss plot of CNN during training

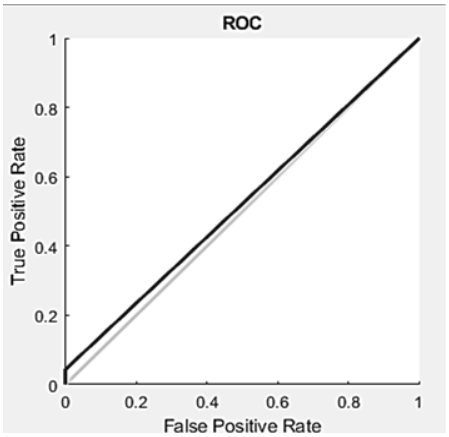


Figure 5. Adaboost plot

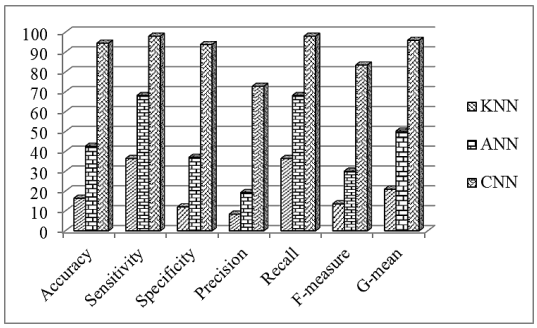


Figure 6. Comparative result of different classifiers

Table 1. Confusion matrix of CNN

		Target class							
		1	2	3	4	5	6	7	
Output class	1	45	0	1	0	0	0	0	97.8%
		13.1%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	2.2%
	2	1	47	0	0	0	0	1	95.9%
		0.3%	13.7%	0.0%	0.0%	0.0%	0.0%	0.3%	4.1%
	3	1	1	47	0	1	0	0	94.0%
		0.3%	0.3%	13.7%	0.0%	0.3%	0.0%	0.0%	6.0%
	4	0	0	0	46	0	1	0	97.9%
		0.0%	0.0%	0.0%	13.4%	0.0%	0.3%	0.0%	2.1%
	5	0	1	0	0	48	0	1	96.0%
		0.0%	0.3%	0.0%	0.0%	14.0%	0.0%	0.3%	4.0%
	6	0	0	1	3	0	48	0	92.3%
		0.0%	0.0%	0.3%	0.9%	0.0%	14.0%	0.0%	7.7%
	7	2	0	0	0	0	0	47	95.9%
		0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	13.7%	4.1%
		91.8%	95.9%	95.9%	93.9%	98.0%	98.0%	95.9%	95.6%
		8.2%	4.1%	4.1%	6.1%	2.0%	2.0%	4.1%	4.4%

Table 2. Real time output results of different classifiers

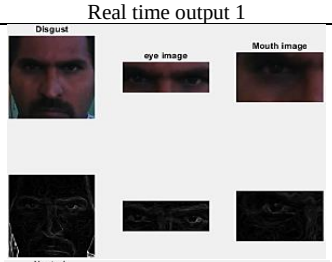
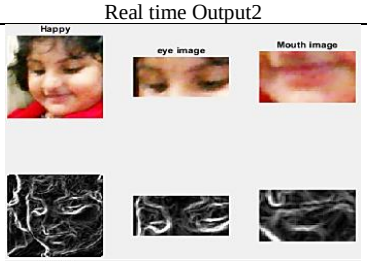
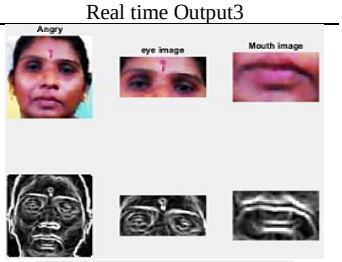
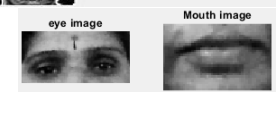
Classifier	Real time output 1	Real time Output2	Real time Output3
ANN			
KNN			
CNN			

Table 3. Performance analysis of different classifiers

Performance metrics (%)	KNN	ANN	CNN
Accuracy	16.39	42.6	94.46
Sensitivity	36.36	68.18	97.96
Specificity	12	37	93.8
Precision	8.33	19.23	72.73
Recall	36.36	68.18	97.96
F-measure	13.56	30	83.48
G-mean	20.89	50.21	95.9

4. CONCLUSION

The performance evaluation of the proposed modified V-J algorithm is carried out with suitable datasets to find the facial emotions from the realtime data-image and also to categorize different emotions. For FER, LPB, GLCM, and RPCA based feature extraction techniques are applied to extract details from face images for reognizing each facial emotion. The entire system is trained and classified using CNN classifiers for FER. The performance of the proposed approach is estimated through the parameters like specificity, sensitivity, precision, recall, and accuracy. The results obtained show that the proposed method efficiently detects emotions in the face images using CNN with an accuracy of 95.33% for different input images.

REFERENCES

[1] T. Chernigovskaya, P. Eismont, and T. Petrova, *Language, music and gesture: informational crossroads*. Springer Singapore, 2021.

[2] S. Mekruksavanich and A. Jitpattanakul, "Biometric user identification based on human activity recognition using wearable sensors: an experiment using deep learning models," *Electronics*, vol. 10, no. 3, p. 308, Jan. 2021, doi: 10.3390/electronics10030308.

[3] A. Swaminathan, A. Vadivel, and M. Arock, "FERCE: facial expression recognition for combined emotions using FERCE algorith," *IETE Journal of Research*, pp. 1–16, May 2020, doi: 10.1080/03772063.2020.1756471.

[4] K. S. Yadav and J. Singha, "Facial expression recognition using modified viola-john's algorithm and KNN classifier," *Multimedia Tools and Applications*, vol. 79, no. 19–20, pp. 13089–13107, May 2020, doi: 10.1007/s11042-019-08443-x.

[5] A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," in *2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, pp. 1–5, doi: 10.1109/incet49848.2020.9154121.

[6] S. K. Mondal, I. Mukhopadhyay, and S. Dutta, "Review and comparison of face detection techniques," in *Proceedings of International Ethical Hacking Conference 2019*, Springer Singapore, 2019, pp. 3–14.

[7] R. Goel, I. Mehmood, and H. Ugail, "A study of deep learning-based face recognition models for sibling identification," *Sensors*, vol. 21, no. 15, p. 5068, Jul. 2021, doi: 10.3390/s21155068.

[8] V. Sreenivas, V. Namdeo, and E. V. Kumar, "Group based emotion recognition from video sequence with hybrid optimization based recurrent fuzzy neural network," *Journal of Big Data*, vol. 7, no. 56, Aug. 2020, doi: 10.1186/s40537-020-00326-5.

[9] L. B. Krithika and G. G. L. Priya, "Graph based feature extraction and hybrid classification approach for facial expression recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2131–2147, Jul. 2020, doi: 10.1007/s12652-020-02311-5.

[10] K. D. Ismael and S. Irina, "Face recognition using viola-jones depending on python," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1513–1521, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1513-1521.

[11] B. Taha and D. Hatzinakos, "Emotion recognition from 2D facial expressions," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, May 2019, pp. 1–4, doi: 10.1109/ccece.2019.8861751.

[12] M. Li, X. Yu, K. H. Ryu, S. Lee, and N. Theera-Umpon, "Face recognition technology development with Gabor, PCA and SVM methodology under illumination normalization condition," *Cluster Computing*, vol. 21, no. 1, pp. 1117–1126, Mar. 2017, doi: 10.1007/s10586-017-0806-7.

[13] M. Nehru and S. Padmavathi, "Illumination invariant face detection using viola jones algorithm," in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Jan. 2017, pp. 1–4, doi: 10.1109/icaccs.2017.8014571.

[14] K. Dang and S. Sharma, "Review and comparison of face detection algorithms," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Jan. 2017, pp. 629–633, doi: 10.1109/confluence.2017.7943228.

[15] L. Shen, H. Wang, L. Da Xu, X. Ma, S. Chaudhry, and W. He, "Identity management based on PCA and SVM," *Information Systems Frontiers*, vol. 18, no. 4, pp. 711–716, Apr. 2015, doi: 10.1007/s10796-015-9551-8.

[16] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," *arXiv preprint*, 2017, doi: 10.48550/arXiv.1703.04691.

[17] V. K. Gudipati, O. R. Barman, M. Gaffoor, Harshagandha, and A. Abuzneid, "Efficient facial expression recognition using adaboost and haar cascade classifiers," in *2016 Annual Connecticut Conference on Industrial Electronics, Technology & Automation (CT-IETA)*, Oct. 2016, pp. 1–4, doi: 10.1109/ct-ieta.2016.7868250.

[18] D. Dagar, A. Hudait, H. K. Tripathy, and M. N. Das, "Automatic emotion detection model from facial expression," in *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, May 2016, pp. 77–85, doi: 10.1109/icaccct.2016.7831605.

[19] A. Garg and R. Bajaj, "Facial expression recognition & classification using hybridization of ICA, GA, and neural network for human-computer interaction," *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 2, no. 1, pp. 49–57, 2015.

[20] N. Mahajan and H. Mahajan, "Emotion detection algorithm," *International Journal of Electrical and Electronics Research*, vol. 2, no. 2, pp. 56–60, 2014.

[21] P. Yaffe, "The 7% rule: fact, fiction, or misunderstanding," *Ubiquity*, vol. 2011, no. October, pp. 1–5, Oct. 2011, doi: 10.1145/2043155.2043156.

[22] K. Nozaki, H. Ishibuchi, and H. Tanaka, "Adaptive fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 238–250, 1996, doi: 10.1109/91.531768.

[23] I. Paliy, "Face detection using Haar-like features cascade and convolutional neural network," in *international conference on" modern problems of radio engineering, telecommunications and computer science"(TCSET)*, 2008, pp. 375–377.

[24] G. UKharat and S. V. D. Ul, "Emotion recognition from facial expression using neural networks," in *2008 Conference on Human System Interactions*, May 2008, pp. 422–427, doi: 10.1109/hsi.2008.4581476.

[25] T. Abidin and W. Perrizo, "SMART-TV: a fast and scalable nearest neighbor based classifier for data mining," in *Proceedings of the 2006 ACM symposium on Applied computing - SAC '06*, 2006, pp. 536–540, doi: 10.1145/1141277.1141403.

BIOGRAPHIES OF AUTHORS

Mrs. Komala Karilingappa obtained B.E in Electronics and Communication Engineering and M.Tech. in Digital Electronics and Communication from Visvesvaraya Technological University, Belagavi in the year 2000 and 2010 respectively. Currently she is working as Assistant Professor, Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India. She has more than 17 years of teaching experience and she published 8 papers in National and Intrernational Journals and conferences, her area of research interest is image processing. She can be contacted at email: komalak@ssit.edu.in.

Devappa Jayadevappa received BE Degree in Instrumentation Technology from Siddaganga Institute of Technology, Tumkur, M.Tech. Degree from SJCE, Mysore specialization in Biomedical Instrumentation and Ph.D. from Jawaharlal Nehru Technological University, Andrapradesh. He is currently working as a Professor, Department of Electronics and Instrumentation Engineering, Jagadguru Sri Shivarathreeshwara Academy of Technical Education, Bengaluru. He has more than 22 years of teaching and industrial experience. He has published more than 100 papers in International and National Journals and Conferences. He is the reviewer for various National and International Journals published across the world. His areas of interests are Digital image processing, Medical imaging, Biomedical Signal Processing and Industrial Automation. He can be contacted at email: djayadevappa@jssateb.ac.in.

Shivaprakash Ganganna received B.E. degree in Instrumentation Technology from SIT, Tumakuru, Karnataka, India, in 1991. He has received a M.Tech. in Bio-medical instrumentation from SJCE, Mysuru, India in 1995. He has received a Ph.D. from Visveswaraya Technological University, Belagaum. He is currently working as an Associate Professor with the Department of Electronics and Instrumentation Engineering, Ramaiah Institute of Technology, Bangalore, Karnataka, India. His research area includes VLSI design, image processing, signal processing, and industrial automation. He can be contacted at email: shivaprakash@msrit.edu.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 87~95
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp87-95

A deep learning based stereo matching model for autonomous vehicle

Deepa[1], Jyothi Kupparu[2]
[1]Department of Information Science and Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology-Affiliated to Nitte (Deemed to be University), Nitte, India
[2]Department of Information Science and Engineering, Jawaharlal Nehru National College of Engineering, Visvesvaraya Technological University, Shimoga, India

Article Info

Article history:

Received Dec 23, 2021
Revised Jul 23, 2022
Accepted Aug 21, 2022

Keywords:

Convolutional neural networks
Disparity
Generative adversarial network
Ill posed regions
Stereo matching

ABSTRACT

Autonomous vehicle is one the prominent area of research in computer vision. In today's AI world, the concept of autonomous vehicles has become popular largely to avoid accidents due to negligence of driver. Perceiving the depth of the surrounding region accurately is a challenging task in autonomous vehicles. Sensors like light detection and ranging can be used for depth estimation but these sensors are expensive. Hence stereo matching is an alternate solution to estimate the depth. The main difficulties observed in stereo matching is to minimize mismatches in the ill-posed regions, like occluded, texture less and discontinuous regions. This paper presents an efficient deep stereo matching technique for estimating disparity map from stereo images in ill-posed regions. The images from Middlebury stereo data set are used to assess the efficacy of the model proposed. The experimental outcome dipicts that the proposed model generates reliable results in the occluded, texture less and discontinuous regions as compared to the existing techniques.

Corresponding Author:

Deepa
Department of Information Science and Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology-Affiliated to Nitte (Deemed to be University)
Nitte, India
Email: deepashetty17@nitte.edu.in

1. INTRODUCTION

Autonomous vehicles are a prominent research topic in the computer vision. It is necessary to correctly measure the three-dimensional (3D) view of the surrounding region of the vehicle in real time to make a driving decision. Precision of the depth map is crucial for the safety measure of autonomous vehicles. In these vehicles the depth details of the surrounding region are usually extracted using the hardware like light detection and ranging sensors. These sensors are expensive to install and also have certain drawbacks that may lower the standard of the depth information. These sensors do not provide additional information like traffic light color which plays a major role in decision making. Computer vision based stereo matching could be an alternate solution to overcome this drawback. The aim of stereo matching is to find matching pixels of images from different viewpoints and then estimate the depth [1]–[3]. It finds its applications in augmented reality, robotics, 3D reconstruction [4]–[9]. Stereo vision tries to imitate the process in the human eye and the human brain. A scene taken from two cameras displaced horizontally will form two slightly separate projections. Disparity is the horizontal displacement in an object. A map that contains displacement of all pixels in an image is known as disparity map. Depth of a scene can be estimated from this disparity map.

In recent past many stereo algorithms were proposed [10]–[12]. A classic stereo algorithm mainly follows three steps namely: computing the pixel wise features, construction of cost volume followed by post-processing. Traditional stereo matching methods are grouped as local, global and semi-global methods. Local methods rely on low level pixel features to compute the similarity in the cost computation step. They estimate the correspondence by means of a window or support region [13]–[15]. Since the pixel wise characterization play a major factor, a wide variety of these representations are used by researchers varying from a simple rgb representation of pixels to the other descriptors like census transform, scale invariant feature transform. Segment based super pixel technique is proposed in [16]. After finding the edges and matching cost, adaptive support weight is used in cost aggregation. It proposes dual path refinement to correct disparities. Stereo matching based on adaptive cross area and guided filtering with orthogonal weights (ACR-GIF-OW) is proposed in [17]. These techniques are computationally less expensive but do not produce accurate results in the texture less, discontinuous and occluded areas.

A Global methods handle texture less regions or uneven surfaces by including smoothness cost. Global methods make use of global energy function. The energy function is minimized step by step to compute disparity by assuming matching as a labelling problem. The pixels are considered as nodes and disparity estimated is considered as labels. The global methods use data and smoothness term to compute the energy function to produce smooth disparity. Graph cut [18], dynamic programming [19] and belief propagation [20] are the classic global matching algorithm. A tree structure is proposed in [21] named pyramid-tree that performs cross regional smoothing and handling region of low texture. In addition, they used log angle for cost computation which is robust to inconsistencies. The performance of global methods is limited because these approaches depend on hand-crafted features and hence do not produce accurate results.

Convolutional neural network (CNN) is popular in different vision [22]–[24] applications. These methods are widely used in stereo matching. It improves the performance as compared to traditional methods. Kendall *et. al.* [25] the authors presented an architecture that learns disparity without regularization. Features are extracted automatically using CNN without any manual intervention. These features are used to perform stereo matching, that can handle texture less regions or uneven surfaces. Eigen *et. al.* [26] made use of basic neural networks to determine depth of a scene. They used AlexNet architecture to generate coarse map. Another network is followed that performs local refinements. The work proposed in [27] included the process of multi-stage framework that combined random forests and CNN. An architecture named neural regression forest is used to find depth from single input image. It allows parallel training of all CNN. Finally, a bilateral filter was used to obtain a refined disparity map. A similar concept is presented in [28] where many tiny neural networks were trained across overlapping patches. DispNet is one of the basic networks used for disparity estimation. A cascading residual learning network is used in [29] that extend the DispNet structure. It is obtained by using DispFullNet and DispResNet. The initial stages of CNN uses DispNet with an additional up convolution module. This help to extract more information. The next stage generates residual signal that helps in refinement. A trainable network is explained in [30]. It uses a robust differentiable patch match internal structure that discards most disparities without performing cost volume evaluation fully. This reduces search space and increases memory and time efficiency. The main drawbacks of existing methods are that the ill posed regions are not handled effectively. In the proposed method CNN is combined with optimization technique. CNN is used to replace the the hand-crafted term with the learned features. The output of CNN is used to calculate the unary and smoothness cost. Smoothness cost is added by taking the information from the neighboring pixels. Smoothness cost estimates the contrast-sensitive information to get a smooth disparity map. Post processing is performed to handle occlusion.

In stereo vision, the areas visible in one view may not be visible in another. It is often difficult to reconstruct such regions in one image by looking at the other. The losses computed in these areas are noisy, leading to inaccurate results specifically in the occluded areas. Disparity refinement is implemented to enhance the accuracy of matching in ill posed areas. The left-right consistency check is the common method used to identify and handle the outliers. Even though several methods were proposed in the past to enhance the efficiency of matching, the low accuracy problem especially in the ill posed areas has not been handled very well. In order to handle these areas, post processing is performed by means of a generative adversarial network (GAN) model put forward by Goodfellow [31]. GAN is a structure used for training generative model. It uses the concept of min-max game. The two models namely generator model and a discriminative model is used to analyze the distribution of data. The generator tries to understand the distribution which is almost same to the real distribution of data. The ability to generate high quality image by GAN makes it applicable in several image processing applications. An encoder decoder structure is used for training in reconstructing the images. This model can produce various realistic representation of input by altering the attribute values. A conditional adversarial network [32] can be used for image translation. This translation converts the image from one representation to the other such as day to night.

We propose a hybrid CNN based deep stereo network model (CDSN) to estimate the disparity map that can produce accurate results. Loopy belief propagation is used to compute initial disparity map from features extracted from CNN. A generative neural network is used to handle the ill posed regions in the disparity map. The generated images look more realistic and closer to ground truth disparity map. The obtained result show that the proposed CDSN model handle the ill posed regions like discontinuities in the image boundaries and occluded areas effectively. The proposed model outperforms the other existing techniques on Middlebury dataset [33]. The paper is organized in a manner, section 2 explains the proposed CDSN model. Section 3 depicts the results of proposed model. The conclusions of the paper are presented in section 4.

2. METHOD

A CNN based model is proposed for stereo matching to find disparity map. The features extracted from CNN is used to compute the unary cost and smoothness cost. Global energy function is adapted to get the initial disparity map. A GAN model is used to handle ill posed region. Table 1 depicts the list of symbols with its description. The flow chart of proposed model is displayed in Figure 1.

Table 1. Symbol table

Symbol	Description
$D(d_i)$	Unary cost at pixel 'i'
V_i^l	Left feature vector
V_i^r	Right feature vector
$S(d_i,d_j)$	Smoothness cost
α and β	Smoothness constants
$msg^t{}_{i\to j}(d_i)$	Message from pixel 'i' to 'j' at iterations 't'
D(p, q)	discriminator
$E_{p,q}$	Expected values of all real data instances
G(p,r)	generator
$E_{p,r}$	Expected values of all generated instances
D_g	Ground truth disparity map
D_t	Estimated disparity map
T	Threshold value
N	Total number of pixels

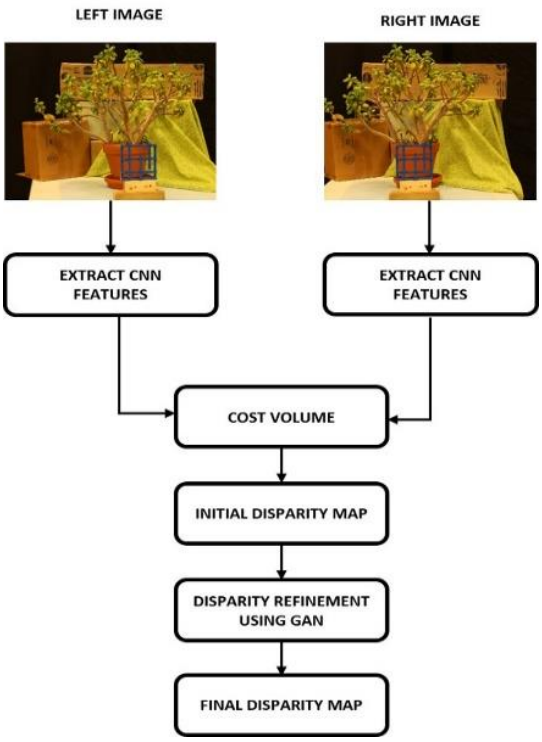


Figure 1. Flowchart of the CDSN model

2.1. CNN feature extraction

Conventional algorithms for stereo matching focuses on hand crafted features which leads to inadequate image information. CNN is used for the various vision problems including stereo matching. The CNN can extract local context better, hence it is robust to any photometric differences. The feature descriptors are extracted from rectified stereo images using a pre-trained visual geometry group (VGG-16) model [34]. The VGG-16 model is trained using ImageNet dataset which contains 14 million labelled images that are of high resolution that belong to 1,000 classes. The output of the 9th layer is used for stereo matching in the proposed model as it presents an appropriate feature space for computing disparity. VGG-16 uses a max pool layer that select the maximum element from the input map using a filter of 2×2. The first and second layers include 64 channels of 3×3 kernel size which is followed by max pool function of stride 2, 2. The third and fourth layers include 128 channels of 3×3 kernel followed by max pool function of stride 2, 2. The next three layers include 256 channels of 3×3 kernel that is followed by max pool function of stride 2, 2. Eighth and ninth layers include 512 channels of 3×3 kernel size. An N-dimensional feature vector is obtained for every location of pixel.

2.2. Initial disparity map estimation

The extracted feature descriptors are used to determine the matching cost of every pixel in left feature map. We search horizontally along the right feature map for the best matching value. The matching unary cost is calculated using the Euclidian distance of two feature descriptors using (1).

$$D(d_i) = min\|V^l - V^r\| \tag{1}$$

Unary cost may not yield optimal result in the texture less, repetitive patterns, discontinuity regions. The smoothness cost is used to smoothen the unary cost. Many smoothening techniques is proposed in the recent past. Most of these methods use random variables to have the disparity of a pixel, which encodes smoothness cost based on some standard constant. The smoothness cost is estimated based on neighbouring pixel information. The smoothness cost penalizes the inconsistent disparity values. The smoothness cost is computed using (2),

$$S(d_i, d_j) = \frac{\propto * (d_i - d_j)^2}{(d_i - d_j)^2 + \beta} \tag{2}$$

Let P represent pixels in the image. The initial disparity map d_i of each pixel $i \in P$ is estimated using energy function E

$$E(d) = \sum_{i \in P} D(d_i) + \sum_{(i,j) \in N} S(d_i, d_j) \tag{3}$$

The proposed method uses max product variation of loopy belief propagation (LBP) [20] to obtain the best disparity map. LBP is an algorithm based on assigning label to each pixel imposing global constraints and message passing. This is an iterative method where the messages are passed to left, right, top and bottom in each iteration. In each iteration t, the message is passing from pixel i to pixel j using (4),

$$msg_{i \to j}^{t}(d_i) = min_{d_i}\left[D(d_i) + S(d_i, d_j) + \sum_{a \in N(i) \setminus j} msg_{a \to i}^{t-1}(d_i)\right] \tag{4}$$

Here a represents all neighbours of i except j
Belief is calculated by (5).

$$Belief(d_i) = D(d_i) + \sum_{k \in N(i)} msg_{k \to i}^{T}(d_i) \tag{5}$$

The values d ranges from 0 to maximum disparity range and k represent neighbours of pixel i . The smooth disparity is obtained for iteration T that minimizes the $Belief(d_i)$. It is observed that the minimization of energy became constant after 10 iterations. Hence the proposed algorithm used 10 iterations.

2.3. Disparity refinement using GAN

The GAN network is used to refine the disparity. This refinement model is used to handle ill posed regions. The GAN can perform learning task automatically by identifying various patterns or irregularities from the input data. GANs have the ability to handle missing data such as occluded pixels in the disparity map. The two sub models in GAN are generator and discriminator. The generator model generates new

samples and discriminator model checks if the generated samples are similar to ground truth map. In the proposed model the network learns through ground truth. The architecture of this refinement technique is given in Figure 2.

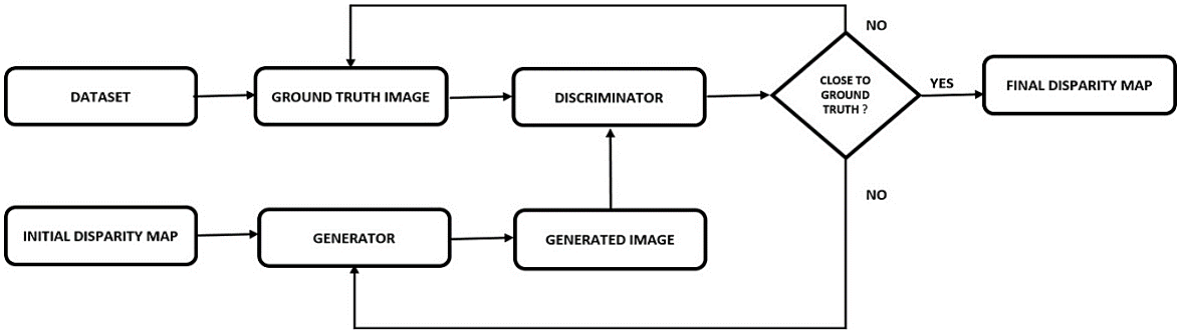


Figure 2. Architecture of disparity refinement technique

The proposed model uses Pix2Pix GAN model [32]. Pix2Pix GAN is simple and can produce high quality images for image translation applications. The efficiency of this GAN as compared to other GAN like CycleGAN [35] and DualGAN [36] is explained in the ablation study. The generator in Pix2Pix is a convolutional network that accepts initial disparity map as the input image and passes it through several convolution and up-sampling layers. Finally, it produces a refined disparity map, where all the occluded areas are filled with valid data. The U-Net auto encoding generator model is trained using adversarial loss that encourages it to create reasonable image. The encoder and decoder are made up of blocks of convolutional, activation layers and batch normalization layers. The generator is updated by loss that is generated between generated image and ground truth image. This information helps generator model to create more reasonable image that is similar to ground truth. The generator G is trained so as to generate output which can be differentiated from ground truth image by a discriminator D. The GAN objective is represented as,

$$L_{GAN}(G,D) = E_{p,q}\left[logD(p,q)\right] + E_{p,r}\left[log\left(1 - D\left(p, G(p,r)\right)\right)\right] \quad (6)$$

Here p denote a ground truth image, q represent the generated image and r represent the initial disparity map

$$G^* = argmin_g max_d L_{GAN}(G,D) \quad (7)$$

G aims to decrease the objective and D aims to increase the objective.
The generator G tries to move the generated image closer to ground truth image using loss L_1 which is calculated as

$$Loss_{L_1}(G) = E_{p,q,r}\left[\left\|\left(q - G(p,r)\right)\right\|_1\right] \quad (8)$$

The final objective is represented as

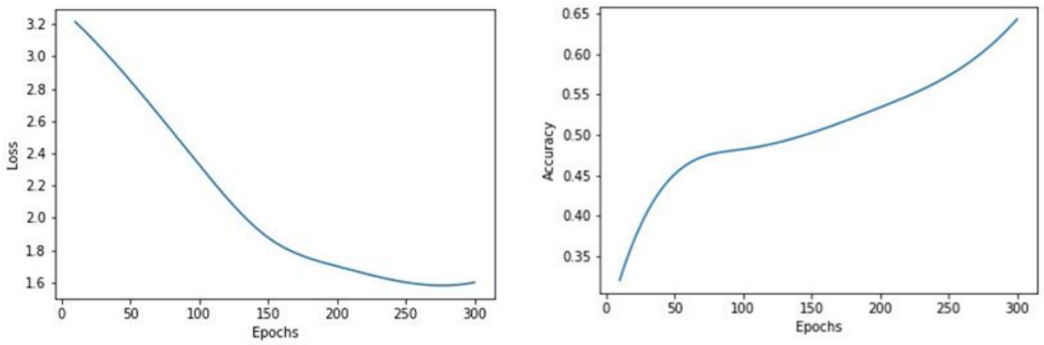
$$G^* = argmin_g max_d L_{GAN}(G,D) + \lambda Loss_{L_1}(G) \quad (9)$$

The visual arti-facts were reduced for the value of λ=100.

The network is trained by images from the Middlebury dataset [33]. The network is tested for 100, 200, 300, 400 epochs. The best disparity map is achieved for 300 epochs. The output from the generator is fed to the discriminator together with ground truth image. The gradient loss is calculated with respect to generator and discriminator to update the model. The trained model is tested to yield a best disparity map. It is observed from the results that best disparity map was obtained by handling the ill posed regions. Figure 3 shows the performace of the model with respect to training loss and training accuracy. Figure 3(a) dipicts training loss and training accuracy against the number of epochs is shown in Figure 3(b). Lower the loss better is the accuracy.

To measure the efficacy of the model proposed, we deployed and tested our model on Dual Intel Xeon E5-2609V4 8C 1.7 GHz 20M 6.4 GT/s with 128GB memory, Dual NVDIA Tesla P100 graphics

processing unit (GPU) with 3584 cores and maximum of 18.7 TeraFLOPS. The proposed CDSN model is evaluated on Middlebury dataset images. These images are pre-processed and rectified stereo images. The output of the 9th layer pre-trained VGG-16 architecture is used for estimating initial disparity map using loopy belief propagation. Initial disparity map is estimated using python programming. GAN is implemented using Pytorch. The Adam optimizer is used to train the Pix2Pix GAN for 300 epochs to handle the ill posed regions. The learning rate has been initialised to 0.0002. The complexity of GAN model is summarized in the Table 2.



(a) (b)

Figure 3. Performance of the model (a) training loss versus number of epochs and (b) training accuracy versus number of epochs

Table 2. Complexity of GAN model

Input size	Optimizer	Parameters	Epochs	Output size	GPU memory	GPU model
256X256X3	Adam	54.414M	300	256X256X3	128GB	Dual NVDIA Tesla P100

3. RESULTS AND DISCUSSION

The proposed model is analyzed for the images taken from Middlebury datasets namely "Jade plant", "Piano", "Pipes", and "Recycle". The test images with resolution are shown in Table 3. Middlebury 2014 dataset contains 33 scenes that are classified into training, additional images and test images. Certain images are used more than once under various exposure. A very high-resolution images is the salient feature of the dataset. Ground truth maps and images are given at quarter, half and full resolution.

Table 3. Images from Middlebury 2014

Images	Image resolution
Jade plant	659x497
Piano	707x481
Pipes	735X485
Recycle	720x486

3.1. Qualitative comparison

The qualitative results for estimating disparity map is depicted in Figure 4. From the top to bottom: Jade plant, piano, pipes, and recycle. Figure 4(a) shows the left image, Figure 4(b) shows the right image, Figure 4(c) represent the ground truth image and Figure 4(d) represent the estimated disparity map.

3.2. Quantitative comparison

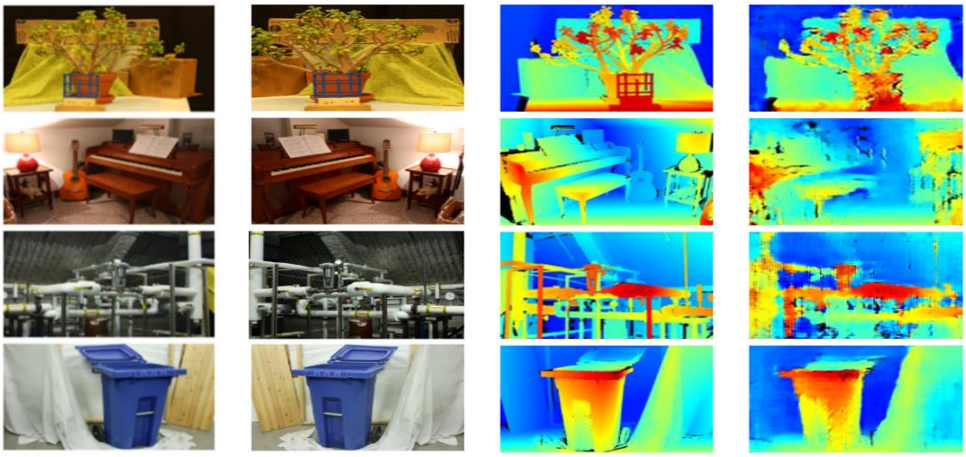
The percentage of bad matching pixel (PBMP) and root mean square error (RMSE) metrics were used for quantitative analysis. Lower values of PBMP and RMSE indicates better efficiency. PBMP is calculated,

$$PBMP = \left[\frac{1}{N}\sum\left|d_t(x,y) - d_g(x,y)\right| > T\right] * 100 \quad (10)$$

RMSE is calculated as,

$$RMSE = \left[\frac{1}{N}\sum\left|d_t(x,y) - d_g(x,y)\right|^2\right]^{\frac{1}{2}} \quad (11)$$

For evaluations purpose, we compared CDSN model with existing stereo matching model. The compared matching models are: deep pruner [30] ACR-GIF-OW [17], and efficient stereo matching by log-angle and pyramid-tree (LPSM) [21]. The occluded areas are not dealt efficiently in [30]. Stereo matching proposed in [17] is computationally less expensive but do not produce accurate results in the texture less, discontinuous areas. Stereo matching proposed in [21] rely on hand-crafted cost matching and hence results produced are not accurate. The Middlebury evaluation leader board results of existing methods are used for comparison. The PBMP and RMSE results of the proposed model and existing techniques are shown in Table 4 and Table 5 respectively. The PBMP and average RMSE results of the proposed CDSN model is less than all three compared method. Hence the proposed model outperforms the compared method and hence suitable for disparity map estimation.



(a) (b) (c) (d)

Figure 4. Visual results on Middlebury images (a) left image, (b) right image, (c) ground truth image and (d) estimated disparity map

Table 4. The quantitative results based on PBMP for error threshold = 1 between computed and ground truth disparities

	Jade plant	Piano	Pipes	Recycle
DEEP PRUNER [30]	62.8	41.0	53.8	36.8
ACR-GIF-OW [17]	51.8	45.1	40.5	37.5
LPSM [21]	59.2	44.8	46.3	36.8
CDSN	50.76	39.71	40.47	33.57

Table 5. The quantitative results based on RMSE between computed and ground truth disparities

	Jade plant	Piano	Pipes	Recycle	Average
DEEP PRUNER [25]	28.2	4.64	13.7	3.81	12.58
ACR-GIF-OW [17]	64.9	14.8	28.6	15.8	31.02
LPSM [21]	34.8	6.09	16.3	5.79	15.74
CDSN	6.07	8.73	8.88	8.48	8.04

3.3. Ablation study

We executed ablation study by comparing the proposed model with the models like CycleGAN and DualGAN. CycleGAN is a technique that performs image translation without using paired examples. This GAN uses unsupervised training. DualGAN is made up of two generators and two discriminators. It is trained to translate images from source to target and target to source. The various metric used are absolute relative distance (ARD), squared relative difference (SRD) and RMSE. Lower values indicate better performance. We find the efficicency of the proposed model is significantly high which is presented in the Table 6.

$$ARD = \frac{1}{N}\sum \frac{d_t(x,y) - d_g(x,y)}{d_t(x,y)} \quad (12)$$

$$SRD = \frac{1}{N}\sum \frac{\left|d_t(x,y) - d_g(x,y)\right|^2}{d_t(x,y)} \quad (13)$$

Table 6. Ablation study using metrics ARD, SRD, RMSE

	CycleGAN	DualGAN	Proposed model	
ARD	0.032	0.035	0.016	Lower is better
SRD	0.352	0.374	0.337	
RMSE	7.036	7.974	6.591	

4. CONCLUSION

This paper presents a novel CNN based model for stereo matching to estimate disparity map from rectified stereo images which is useful in autonomous vehicles. The features extracted from CNN is used to compute the unary cost and smoothness cost. The initial disparity map is obtained using loopy belief propagation, which is then refined using a GAN model to handle the ill posed regions. It is found that the proposed model based on CNN generated disparity maps which are smoother than those generated using naive model and the ill posed regions are handled well using GAN network. The proposed model is evaluated qualitatively as well as quantitatively on various images from Middlebury stereo data set. The results determine that proposed model achieves best disparity map and outperforms existing methods.

REFERENCES

[1] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, no. 1, pp. 7–42, 2002, doi: 10.1109/ SMBV.2001.988771.
[2] R. A. Hamzah and H. Ibrahim, "Literature survey on stereo vision disparity map algorithms," *Journal of Sensors*, vol. 2016, 2016, doi: 10.1155/2016/8742920.
[3] M. S. Hamid, N. F. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Stereo matching algorithm based on deep learning: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1663–1673, 2022, doi: 10.1016/j.jksuci.2020.08.011.
[4] U. Hani and L. Moin, "Realtime autonomous navigation in V-Rep based static and dynamic environment using EKF-SLAM," *IAES International Journal of Robotics and Automation (IJRA)*, vol. 10, no. 4, p. 296, 2021, doi: 10.11591/ijra.v10i4.pp296-307.
[5] Susanto, D. D. Budiarjo, A. Hendrawan, and P. T. Pungkasanti, "The implementation of intelligent systems in automating vehicle detection on the road," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 571–575, 2021, doi: 10.11591/ijai.v10.i3.pp571-575.
[6] S. Sivaraman and M. M. Trivedi, "A review of recent developments in vision-based vehicle detection," *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 310–315, 2013, doi: 10.1109/IVS.2013.6629487.
[7] S. Hong, M. Li, M. Liao, and P. Van Beek, "Real-time mobile robot navigation based on stereo vision and low-cost GPS," *IS and T International Symposium on Electronic Imaging Science and Technology*, pp. 10–15, 2017, doi: 10.2352/ISSN.2470-1173.2017.9.IRIACV-259.
[8] B. Krajancich, P. Kellnhofer, and G. Wetzstein, "Optimizing depth perception in virtual and augmented reality through gaze-contingent stereo rendering," *ACM Transactions on Graphics*, vol. 39, no. 6, 2020, doi: 10.1145/3414685.3417820.
[9] H. Ham, J. Wesley, and Hendra, "Computer vision based 3D reconstruction : a review," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, pp. 2394–2402, 2019, doi: 10.11591/ijece.v9i4.pp2394-2402.
[10] C. Bai, Q. Ma, P. Hao, Z. Liu, and J. Zhang, "Improving stereo matching algorithm with adaptive cross-scale cost aggregation," *International Journal of Advanced Robotic Systems*, vol. 15, no. 1, 2018, doi: 10.1177/1729881417751544.
[11] H. Shabanian and M. Balasubramanian, "A new hybrid stereo disparity estimation algorithm with guided image filtering-based cost aggregation," *IS and T International Symposium on Electronic Imaging Science and Technology*, vol. 2021, no. 2, 2021, doi: 10.2352/ISSN.2470-1173.2021.2.SDA-059.
[12] R. A. Hamzah, M. G. Y. Wei, and N. S. N. Anwar, "Development of stereo matching algorithm based on sum of absolute RGB color differences and gradient matching," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2375–2382, 2020, doi: 10.11591/ijece.v10i3.pp2375-2382.
[13] H. Liu, R. Wang, Y. Xia, and X. Zhang, "Improved cost computation and adaptive shape guided filter for local stereo matching of low texture stereo images," *Applied Sciences (Switzerland)*, vol. 10, no. 5, 2020, doi: 10.3390/app10051869.
[14] S. Chen, J. Zhang, and M. Jin, "A simplified ICA-based local similarity stereo matching," *Visual Computer*, vol. 37, no. 2, pp. 411–419, 2021, doi: 10.1007/s00371-020-01811-x.
[15] J. K and P. C.J, "Multi modal face recognition using block based curvelet features," *International Journal of Computer Graphics & Animation*, vol. 4, no. 2, pp. 21–37, 2014, doi: 10.5121/ijcga.2014.4203.
[16] C. S. Huang, Y. H. Huang, D. Y. Chan, and J. F. Yang, "Shape-reserved stereo matching with segment-based cost aggregation and dual-path refinement," *Eurasip Journal on Image and Video Processing*, vol. 2020, no. 1, 2020, doi: 10.1186/s13640-020-00525-3.
[17] L. Kong, J. Zhu, and S. Ying, "Local stereo matching using adaptive cross-region-based guided image filtering with orthogonal weights," *Mathematical Problems in Engineering*, vol. 2021, 2021, doi: 10.1155/2021/5556990.
[18] V. Kolmogorov, P. Monasse, and P. Tan, "Kolmogorov and zabih's graph cuts stereo matching algorithm," *Image Processing On Line*, vol. 4, pp. 220–251, 2014, doi: 10.5201/ipol.2014.97.
[19] O. Veksler, "Stereo correspondence by dynamic programming on a tree," *Proceedings - 2005 IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition, CVPR 2005, vol. II, pp. 384–390, 2005, doi: 10.1109/CVPR.2005.334.
[20] J. Sun, H. Y. Shum, and N. N. Zheng, "Stereo matching using belief propagation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2351, pp. 510–524, 2002, doi: 10.1007/3-540-47967-8_34.
[21] C. Xu, C. Wu, D. Qu, F. Xu, H. Sun, and J. Song, "Accurate and efficient stereo matching by log-angle and pyramid-tree," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 4007–4019, 2021, doi: 10.1109/TCSVT.2020.3044891.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.
[23] M. S. Hamid, N. A. Manap, R. A. Hamzah, A. F. Kadmin, S. F. A. Gani, and A. I. Herman, "A new function of stereo matching algorithm based on hybrid convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 1, pp. 223–231, 2022, doi: 10.11591/ijeecs.v25.i1.pp223-231.
[24] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017, doi: 10.1109/TPAMI.2016.2572683.
[25] A. Kendall *et al.*, "End-to-end learning of geometry and context for deep stereo regression," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 66–75, 2017, doi: 10.1109/ICCV.2017.17.
[26] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," *Advances in Neural Information Processing Systems*, vol. 3, no. January, pp. 2366–2374, 2014.
[27] A. Roy and S. Todorovic, "Monocular depth estimation using neural regression forest," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 5506–5514, 2016, doi: 10.1109/CVPR.2016.594.
[28] A. Chakrabarti, J. Shao, and G. Shakhnarovich, "Depth from a single image by harmonizing overcomplete local network predictions," *Advances in Neural Information Processing Systems*, pp. 2666–2674, 2016.
[29] J. Pang, W. Sun, J. S. J. Ren, C. Yang, and Q. Yan, "Cascade residual learning: a two-stage convolutional neural network for stereo matching," *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, vol. 2018-Janua, pp. 878–886, 2017, doi: 10.1109/ICCVW.2017.108.
[30] S. Duggal, S. Wang, W. C. Ma, R. Hu, and R. Urtasun, "Deeppruner: learning efficient stereo matching via differentiable patchmatch," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-Octob, pp. 4383–4392, 2019, doi: 10.1109/ICCV.2019.00448.
[31] I. J. Goodfellow *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 3, no. January, pp. 2672–2680, 2014, doi: 10.3156/jsoft.29.5_177_2.
[32] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5967–5976, 2017, doi: 10.1109/CVPR.2017.632.
[33] D. Scharstein *et al.*, "High-resolution stereo datasets with subpixel-accurate ground truth," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8753, pp. 31–42, 2014, doi: 10.1007/978-3-319-11752-2_3.
[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
[35] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 2242–2251, 2017, doi: 10.1109/ICCV.2017.244.
[36] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: unsupervised dual learning for image-to-image translation," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 2868–2876, 2017, doi: 10.1109/ICCV.2017.310.

BIOGRAPHIES OF AUTHORS

Deepa is a graduate with M. Tech in Computer Science Engineering from N.M.A.M. Institute of Technology, Nitte, India. She is currently pursuing her Ph.D. degree in Computer Science engineering at VTU. She is currently working as an assistant professor at Information Science & Engineering in N.M.A.M. Institute of Technology, Nitte, India. Her research interests are in fields of computer vision, digital image processing. She has published several papers in international journals and conferences. She can be contacted at email: deepashetty17@nitte.edu.in.

Jyothi Kupparu received the Ph. D degree in computer science from the Kuvempu University, Shimoga, India. She is a Professor of Information Science & Engineering at J.N.N.C.E Shimoga. Her research interests include image processing, stereo correspondence algorithms for face images, multimodal face recognition, 3D sparse reconstruction and techniques based on stereo rectification for face images. She has published several papers in international journals and conferences. She can be contacted at email: jyothik@jnnce.ac.in.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 96∼105
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp96-105

Boosting auxiliary task guidance: a probabilistic approach

Irfan Mohammad Al Hasib[1], Sumaiya Saima Sultana[1], Imrad Zulkar Nyeen[2], Muhammad Abdus Sabur[2]
[1]Department of Mechanical Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

Article Info

Article history:

Received Dec 26, 2021
Revised Aug 31, 2022
Accepted Sep 26, 2022

Keywords:

Auxiliary task guidance
Computer vision
Deep learning
Multi task learning
Visual odometry

ABSTRACT

This work aims to introduce a novel approach for auxiliary task guidance (ATG). In this approach, our goal is to achieve effective guidance from a suitable auxiliary task by utilizing the uncertainty in calculated gradients for a mini-batch of samples. Our method calculates a probabilistic fitness factor of the auxiliary task gradient for each of the shared weights to guide the main task at every training step of mini-batch gradient descent. We have shown that this proposed factor incorporates task specific confidence of learning to manipulate ATG in an effective manner. For studying the potency of the method, monocular visual odometry (VO) has been chosen as an application. Substantial experiments have been done on the KITTI VO dataset for solving monocular VO with a simple convolutional neural network (CNN) architecture. Corresponding results show that our ATG method significantly boosts the performance of supervised learning for VO. It also out performs state-of-the-art (SOTA) auxiliary guided methods we applied for VO. The proposed method is able to achieve decent scores (in some cases competitive)compared to existing SOTA supervised monocular VO algorithms, while keeping an exceptionally low parameter space in supervised regime.

Corresponding Author:

Irfan Mohammad Al Hasib
Department of Mechanical Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: irfanhasib.me@gmail.com

1. INTRODUCTION

Recent research shows that various forms of multi-task learning (MTL) are being used to boost the performance of neural networks (NN) beyond their capacity [1]. The goal of MTL is to maximize the performance of several tasks by learning multiple tasks together while auxiliary task learning is directly concerned with better performance of the primary task [2]. In this work, we present an effective approach to gain guidance from auxiliary tasks. The novelty of the proposed approach is that it can effectively control the contribution of auxiliary task gradients for each shared weight by measuring its suitability for fitting into the main task's loss gradient distribution. For investigating the effectiveness of the proposed method, we have chosen the complex problem of monocular visual odometry (VO) pose estimation. Geometric approaches of VO [3]–[6] work very well for known environments, but require consistency in camera calibrations [7]. However, learning based approaches show superiority in robustness to inconsistent environment[8]. Complex architectures of deep learning (DL) solutions capture the high complexity of the VO problem. However, DL based approaches possess limitations like higher inference time, larger memory requirements, and overfitting tendencies. Simpler architectures may create balance between these challenges. But merely using a simple architecture is not good enough for solving complex VO problems [9]. Performance boosting techniques like MTL, auxiliary task learning (ATL) can be a solution here. Costante and Ciarfuglia [10], Yang *et al.* [11] are examples where MTL

approaches have been embraced in pose estimation. Using the proposed ATG method, we solve them monocular VO pose estimation successfully problem with relatively simple architecture.

Developing better guidance methods for MTL and ATL is a primary research question. Chen *et al.* [12] performs normalization of gradients to balance learning between multiple tasks. Yu *et al.* [13] manipulates directions of the gradients to provide better guidance. Du *et al.* [14] quantifies similarity between this by measuring the cosine similarity between gradient vectors of two tasks and therefore tuning for a suitable threshold for similarity value. Our proposed approach measures similarity in a more precise manner by considering each shared weight gradient separately. Unlike existing approaches, we weigh the similarity with task specific confidence of learning as well. For the chosen field of application of VO, traditional geometric methods produced state-of-the-art solutions for pose estimation including [6], [5], but they are prone to motion drift. Supervised learning-based methods solve this challenge because they are more robust to unstable environments [15]. However, they possess an additional challenge of requiring complex architectures or having a huge number of hyper-parameters [16]–[21]. Due to these conflicts, ultimately most recent works are focusing on unsupervised learning and being successful with much higher margin [11], [22], [23]. Among MTL based supervised approaches for VO problem, latent space VO (LS-VO) [10] learns a low dimensional optical flow (OF) subspace with pose estimation jointly but still works in a huge parameter space. Our proposed ATG approach helps enable a supervised learning method to perform well even in a much lower parameter space than existing methods for the complex problem of VO. At a glance, the contributions of this paper are: i) Proposing a new approach for providing effective and better guidance from auxiliary task. ii) Demonstrating the effectiveness of the proposed method by solving the monocular VO problem using a tiny network compared to existing supervised models with OF subspace learning as auxiliary task.

2. METHODOLOGY

2.1. NN architecture specification

The complete architecture, illustrated in Figure 1 can be divided into two major sections, encoder section and task specific section. The encoder section is a modified FlowNet architecture [24], reducing its depth by half for every layer. This section is the feature extractor of the framework and is shared by both tasks. The task specific section, consisting of three parts, is dedicated for translation estimation, rotation estimation and flow image prediction. Rotation and translation estimation parts of the network are based on separate sequences of dense fully connected layers and a decoder network estimates OF.

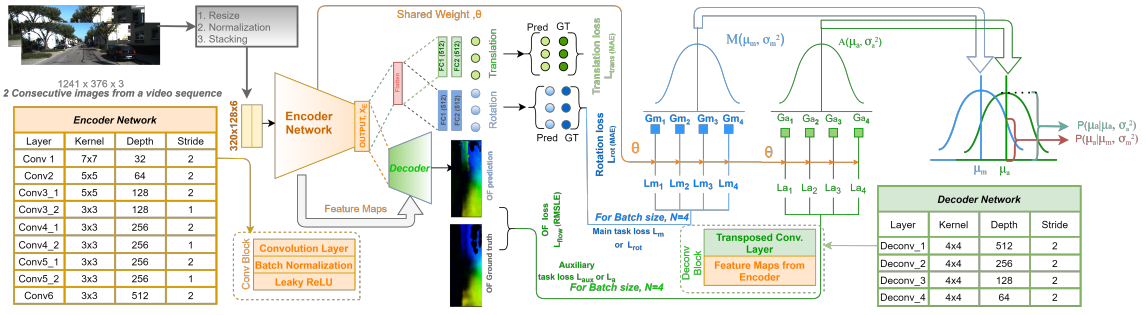


Figure 1. Visual representation of the architecture and algorithm

We scaled down the input images from $(1241, 376, 3)$ to $(320, 128, 3)$ which reduces parameters and helps to overcome overfitting without compromising result accuracy. Images were normalized with mean centered to 0 ranging from -0.5 to +0.5. The network takes two consecutive images from each particular sequence of KITTI VO dataset and stacks them depth-wise as input. In Figure 1, output $\mathbf{X_E}$ is generated after the image passes through the 9 shared layers of the encoder. Each shared layer block consists of 2D-convolution layer, batch normalization and leaky rectified linear unit (leaky ReLU). The flow image prediction section uses output $\mathbf{X_E}$ directly as the input. Flattened $\mathbf{X_E}$ is used as the input of translation estimation and rotation estimation sections, both consisting of dense layers. Final outputs of the entire framework are 6 degrees of freedom (DOF) pose estimation and flow image predictions.

2.2. Probabilistic auxiliary task guidance

As shown in Figure 1, the network learns three tasks with three different loss functions-translation loss(L_{trans}), rotation loss(L_{rot}) (as main task loss), OF subspace learning loss(L_{flow}) (as auxiliary task loss). In ATG learning, the total loss is usually defined as:

$$L_{total}(\theta, \phi_{main}, \phi_{aux}) = \beta_m L_{main}(\theta, \phi_{main}) + \beta_a L_{aux}(\theta, \phi_{aux})$$
$$\text{weight } \theta \text{ can be updated as, } \theta_{new} = \theta + \alpha(\beta_m \frac{1}{N} \sum_{i=1}^{N} \frac{dL_{main}}{d\theta} + \beta_a \frac{1}{N} \sum_{i=1}^{N} \frac{dL_{aux}}{d\theta})$$

Here, α= learning rate, β_m= main task loss coefficient, β_a= auxiliary task loss coefficient, θ= weights of shared layer, ϕ_{main}= weights of task specific layers for main task, ϕ_{aux}= weights of task specific layers for auxiliary task, N= mini-batch size. One of the key research questions in ATL is to choose an optimum coefficient (β_a) to encourage positive transfer and blocking negative transfer from an appropriate auxiliary task [1], [14]. Our approach finds a solution to this question by tuning β_a initially and then optimizing it extensively with a probabilistic factor calculated for each shared weight that prioritizes assistance from the auxiliary task with respect to its guiding capability. In this section, we present the approach of calculating this factor and discuss how it allows us to integrate both task specific confidence of learning and task similarity in the guidance process.

From central limit theorem, we can say that the gradients of a mini-batch belong to a certain normal distribution. Let the mean of the gradients for main task, $\frac{1}{N} \sum_{i=1}^{N} \frac{dL_{main}}{d\theta}$ be μ_m and the mean of the gradients for auxiliary task be μ_a. The distributions of gradients for the main and auxiliary task are denoted as $M(\mu_m, \sigma_m^2)$ and $A(\mu_a, \sigma_a^2)$ respectively. Calculated probabilities of μ_a in both distributions have been expressed as $P(\mu_a|\mu_m, \sigma_m^2)$ and $P(\mu_a|\mu_a, \sigma_a^2)$. This calculated $P(\mu_a|\mu_m, \sigma_m^2)$ indicates what probability would μ_a have if it belonged to the distribution of the main task M. In other words, it signifies how much μ_a fits the current distribution M as a random numeric value. Hence, it could be a reasonable parameter to decide how much auxiliary task loss should contribute to the total loss. It can be incorporated by using it as a multiplication factor with auxiliary task loss coefficient. But empirical values of gradients and their variances reveal that the probability $P(\mu_a|\mu_m, \sigma_m^2)$ values vary in between a very wide range($10^1 \sim 10^4$), shown in Figure 2. Hence, using probability $P(\mu_a|\mu_m, \sigma_m^2)$ merely as the multiplication factor makes the gradients unstable by changing them drastically. Two types of scaling are done to handle this issue. In the first method, we divide these probabilities by their maximum value in respective layer's weights. Thus it is restricted between (0,1). This method is named probabilistic factor (PF) method.

$$\Delta\theta = \beta_m \mu_m + [P(\mu_a|\mu_m, \sigma_m^2)/P_{max}]\beta_a \mu_a \tag{1}$$

Where, P_{max} = maximum value of probability in respective layers. However, probability values with small magnitude are at risk of getting vanished especially when variance values are high. Taking log of probability does not help much in this issue. So,in the second method, we propose to scale the probability $P(\mu_a|\mu_m, \sigma_m^2)$ by dividing it with $P(\mu_a|\mu_a, \sigma_a^2)$ which is the probability of the same variable μ_a in its own distribution. We denote this method as the probability ratio factor (PRF) (1) method. It has performed best in our experiments for reasons we will explain gradually in later sections. We have defined the ratio of two probabilities as relative probability factor, $\rho(m, a)$ and updated (1) as (3).

$$\rho(m, a) = \frac{P(\mu_a|\mu_m, \sigma_m^2)}{P(\mu_a|\mu_a, \sigma_a^2)} \tag{2}$$

$$\Delta\theta = \beta_m \mu_m + \rho(m, a)\beta_a \mu_a \tag{3}$$

$$\text{So, } \theta_{new} = \theta + \alpha(\beta_m \frac{1}{N} \sum_{i=1}^{N} \frac{dL_{main}}{d\theta} + \rho(m, a)\beta_a \frac{1}{N} \sum_{i=1}^{N} \frac{dL_{aux}}{d\theta})$$

The value of $\rho(m, a)$ is constrained within a suitable range (shown in Figure 2(a), 2(b), and 2(c)). So the ratio does not change drastically for consecutive training steps, the learning process becomes more stable. For the purpose of analysis ,we have introduced two novel terms: task confidence, ζ and task similarity, τ as:

$$P(\mu_a|\mu_m, \sigma_m^2) = \frac{1}{\sigma_m\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{\mu_a - \mu_m}{\sigma_m})^2); \zeta(m) = \frac{1}{\sigma_m\sqrt{2\pi}}; \tau(m, a) = \exp(-\frac{1}{2}(\frac{\mu_a - \mu_m}{\sigma_m})^2)$$

The term $\frac{1}{\sigma_m\sqrt{2\pi}}$ is inversely proportional to standard deviation (σ_m) of the distribution. For a particular mini-batch, lower variance of the gradients will indicate stable learning. Intuitively, we interpret it as a measure of confidence of this distribution M.The term $(\frac{\mu_a-\mu_m}{\sigma_m})$ is the measure of distance between μ_a and μ_m for unit σ_m. So, mathematically, $\tau(m,a)$ will measure similarity(not distance) very strictly if the distribution has low variance and vice versa. For auxiliary task, $\tau(a,a) = \exp(-\frac{1}{2}(\frac{\mu_a-\mu_a}{\sigma_a})^2) = 1$. So, from (2), probability ratio,

$$\rho(m,a) = \frac{P(\mu_a|\mu_m,\sigma_m^2)}{P(\mu_a|\mu_a,\sigma_a^2)} = \frac{\zeta(m)\tau(m,a)}{\zeta(a)\tau(a,a)} = \frac{\zeta(m)}{\zeta(a)}\tau(m,a) \tag{4}$$

Here, we have defined $\zeta(m)/\zeta(a)$ as the relative task confidence of the main task compared to the auxiliary task which helps main task in different scenarios, summarized in Table 1. If the main task is learning with relatively higher confidence and auxiliary task gradient still has a good similarity (despite the high confidence value of the main task) that is the most desired scenario for ATG. Thus our approach ensures efficient and effective guidance from the auxiliary task by avoiding negative transfer in critical scenarios.

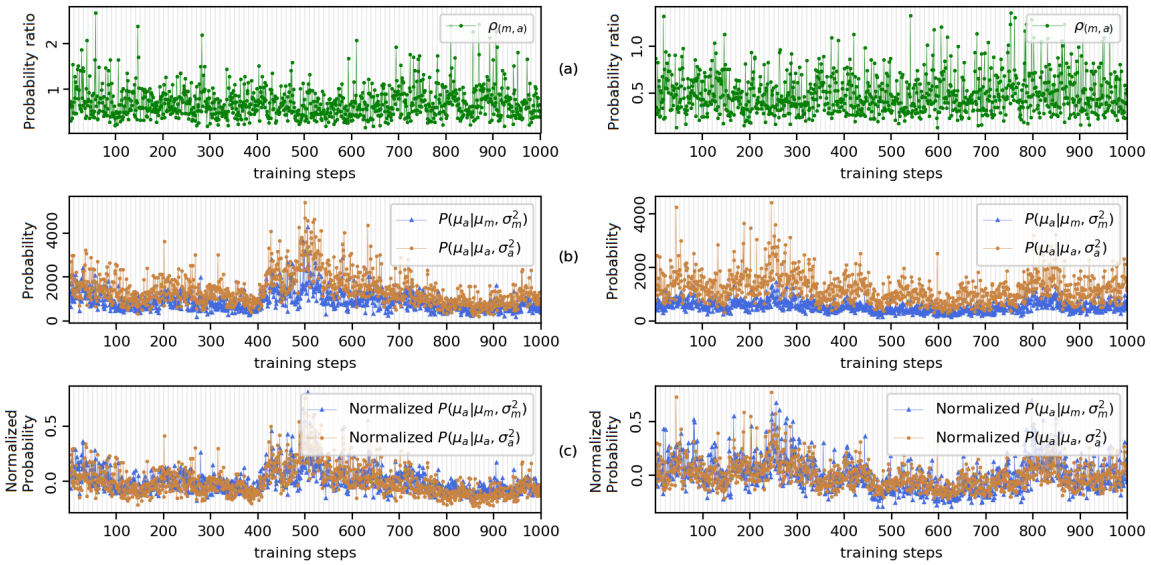


Figure 2. Probability ratio, raw probability and normalized probabilities for main task and auxiliary task. Data is collected for one random weight of layer 1 for 1,000 consecutive training steps. Left plot is for 15^{th} epoch and right plot is for 30^{th} epoch among 30 epochs. a) probability ratio b) raw probabilities and c) normalized probabilities.

Table 1. Possible scenarios of the PRF method

Scenario	$\zeta(m)/\zeta(a)$	$\tau(m,a)$	Auxiliary task guidance
Case 1	high	low	moderate guidance
Case 2	high	high	helps positive guidance
Case 3	low	low	blocks negative guidance
Case 4	low	high	moderate guidance

From the above definition of the given terms, we can write, $P(\mu_a|\mu_m,\sigma_m^2) = \zeta(m)\tau(m,a)$. Now, probabilities vary within a wider range around $(10^1 \sim 10^4)$ [2(b)]. Since task similarity, $\tau(m,a)$ lies between $(0 \sim 1)$, so the value of ζ effectively controls the range of probability values. Consequently correlation between $\zeta(m)$ and $\zeta(a)$ shown in Figure 3 causes correlation between the probabilities demonstrated in Figure 2(c). That's why the scaling effectively keeps the probability ratio $\rho(m,a)$ within a suitable range [Figure 2(a)]. The relative probability factor reduces the effect of common sources of variance among the distributions. Thus it emphasizes on the task specific variance that gives information only about relative performance of the tasks. It will be explained in detail in 2.4.. Also, auxiliary task gradients are more stable compared to the main task,

(discussed in 2.3.). So, the relative probability ($\rho(m,a)$) gives us a better estimate of relative performance of the main task compared to auxiliary task. Note that the parameter β_a is a constant coefficient (tuned as a hyperparameter) for all the weights but $\rho(m,a)$ is calculated separately for each of the weights at every mini-batch gradients update.

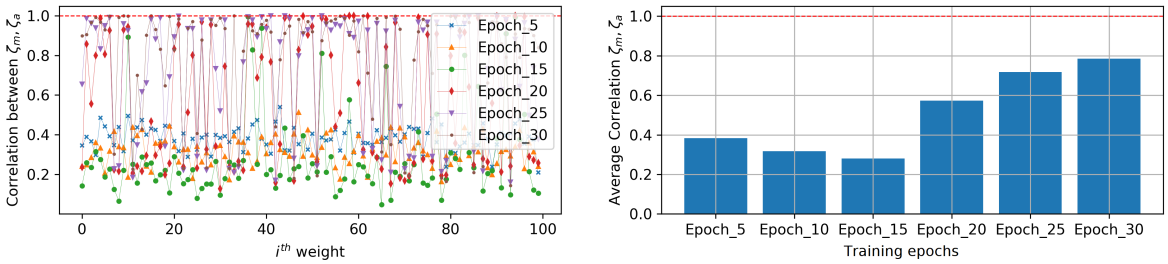


Figure 3. Correlation of $\zeta(a)$ and $\zeta(m)$ for i=100 different weights (θ_i) (X-axis). Here, $\zeta(m)$ and $\zeta(a)$ is collected from 1,000 consecutive steps of a single epoch and $corr(\zeta(m),\zeta(a))$ for each of the i^{th} weight is measured from this 1,000 confidence pairs. To comprehend the information, right side plots are numerical average for every epochs of the left plots

2.3. Optical flow subspace learning as auxiliary task

Estimating a lower dimensional representation of the dense OF field from one pair of raw red green blue (RGB) images is the auxiliary task learned by our network. The chosen auxiliary task is easier to learn than the main task. This is because latent OF representation estimation loss is measured from the entire OF image where the odometry loss is measured from 6 sparse pose values. Also we are learning OF subspace by minimizing pixel-wise root mean squared log error (RMSLE) (while precise raw OF learning requires using root mean squared error (RMSE) loss). This makes the learning task even easier [10]. Finally, the capability of OF task for helping pose estimation can be proved from vanilla MTL results from experiment with ATG method in Table 2.

Table 2. Comparative results of different approaches

Sequences	NN		ATG		PF		PRF	
	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}
04	13.5811	2.0449	12.0222	2.2133	11.7648	0.9098	15.3180	1.1278
05	10.9609	2.1660	11.9858	2.1768	9.4700	1.8930	6.9010	1.4982
07	13.1821	4.5040	11.6105	4.3484	8.0245	3.7180	6.9114	2.5165
10	21.9880	8.0013	19.3150	3.9233	13.5378	4.2081	11.6025	3.4710

t_{rel}: mean translational RMSE drift (%) on length of 100m-800m.
r_{rel}: mean rotational RMSE drift ($\deg/100m$) on length of 100m-800m.

2.4. Sources of variance in gradients

In this section, we have analyzed the sources of variances in the task specific loss gradients calculated with respect to the shared weights. The loss gradients can be expressed in (5) using chain rule of differentiation:

$$G_t = \frac{dl_t}{d\theta} = \frac{dz}{d\theta}.\frac{dA}{dz}.\frac{dl_t}{dA} \tag{5}$$

Where, $z = \theta x + b$, $A = f(z)$, x = input coming from previous layer, l_t = function of loss for task t, θ = shared weight; b = bias; f:activation function. Since the shared layers are convolutional layers the above equations element should be corresponding matrices like $\mathbf{Z} = \Theta\ \mathbf{x} + \mathbf{b}$. But we have considered scaler variables for simplicity. From (5), the gradient equations of both tasks will be the same except $\frac{dl_m}{dA}$ for main task and $\frac{dl_a}{dA}$ for auxiliary task. So we can write $dz/d\theta \times dA/dz = F(x)$ since z and A both are functions of x and $dl_t/dA = H(l_t)$. Consequently, the common term $F(x)$ is responsible for the correlation in the variances of these gradients as shown in Figure 3. Generalizing the equation for gradient calculation:

$$G_t = \frac{dl_t}{d\theta} = F(x)H(l_t) \tag{6}$$

We propose that incorporating relative task confidence $\zeta(m)/\zeta(a)$, allows us to diminish the effect of the part of variance coming from the common source. So the only functional part of the variance is coming from task specific losses. This claim can be proved by following example:
Let $X = \{x_1, x_2, ..., x_D\}$; $L_m = \{l_1^m, l_2^m, ..., l_D^m\}$; $L_a = \{l_1^a, l_2^a, ..., l_D^a\}$ (D is the size of dataset) be a set of inputs and corresponding main task and auxiliary task losses. Let's consider a mini batch of n samples, $X_b = \{x_i | i \in [1, n], i \in \mathbb{N}\}$ where $X_b \subseteq X$; $L_b^m = \{l_i^m | i \in [1, n], i \in \mathbb{N}\}$ where $L_b^m \subseteq L^m$; $L_b^a = \{l_i^a | i \in [1, n]\}$ where $L_b^a \subseteq L_a$. Let, $n = 3$. For ease of expressing relations, elements of set $F(X_b)$ are denoted by a, b, c respectively. Similarly, $H(L_b^m) = \{u, v, w\}$ and $H(L_b^a) = \{p, q, r\}$. So, from (6), gradients can be expressed as:

$$G_1^m = au; G_2^m = bv; G_3^m = cw; G_1^a = ap; G_2^a = bq; G_3^a = cr$$

$$\frac{\zeta(m)}{\zeta(a)} = \frac{\sqrt{E(G_b^{a^2}) - E(G_b^a)^2}}{\sqrt{E(G_b^{m^2}) - E(G_b^m)^2}} = \frac{\sqrt{[N(G_1^{a^2} + G_2^{a^2} + G_3^{a^2}) - (G_1^a + G_2^a + G_3^a)^2]}}{\sqrt{[N(G_1^{m^2} + G_2^{m^2} + G_3^{m^2}) - (G_1^m + G_2^m + G_3^m)^2]}}$$

$$\frac{\zeta(m)}{\zeta(a)} = \frac{\sqrt{[N(a^2p^2 + b^2q^2 + c^2r^2) - (ap + bq + cr)^2]}}{\sqrt{[N(a^2u^2 + b^2v^2 + c^2w^2) - (au + bv + cw)^2]}} \tag{7}$$

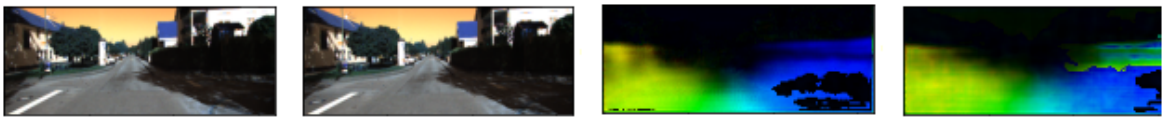
If the loss dependent variable sets (p, q, r in numerator and u, v, w in denominator) have very low variance compared to the input dependent variable set a, b, c; then the change in values of σ_a and σ_m will be dominated by mostly a, b, c. In (7), the denominator and the numerator both contain input dependent variables a, b, c which consequently causes correlation between $\zeta(m)$ and $\zeta(a)$ (observed in Figure 3). The lower the variance of p, q, r and u, v, w will be (compared to the variance of a, b, c), the higher correlation will be eventually. Let, σ_{abc} = standard deviation of the input dependent variables, σ_{uvw} = standard deviation of main task loss dependent variables, σ_{pqr} = standard deviation of auxiliary task loss dependent variables. Let's discuss the effect of σ_{abc} and σ_{uvw} in relative task confidence, $\zeta(m)/\zeta(a)$, considering σ_{pqr} remains almost constant. Confidence of main task $\zeta(m)$ decreases when σ_m increases. This can happen in 3 possible cases: *Case 1*: only σ_{abc} increases- In this case the numerator and denominator both will increase resulting comparatively no notable change in relative task confidence factor,$\zeta(m)/\zeta(a)$. So it diminishes the effect of high variance of main task loss gradient if the variance is being caused by input dependent sources (common source). *Case 2*: σ_{uvw} increases- In this case only the denominator will increase, resulting in significantly lower $\zeta(m)/\zeta(a)$ value. So here $\zeta(m)/\zeta(a)$ is considering the effect of high variance of main task loss gradient significantly only when the variance is being caused by task specific sources (uncommon source which is task loss dependent). *Case 3*: Both σ_{abc} and σ_{uvw} increases- In this case the numerator will increase but the denominator will increase more since it is function of both a, b, c and u, v, w, resulting moderately lower value of $\zeta(m)/\zeta(a)$ (not as low as case 2). So, $\zeta(m)/\zeta(a)$ seems to decrease the effect of high variance of main task loss gradient moderately because the variance is being caused by both task loss specific sources and input dependent sources. Above 3 cases demonstrate, how relative task confidence is less affected by common source of variances.

Algorithm 1 Algorithm (PRF)

$Init\ \theta, \phi_{main}, \phi_{aux}$
$set\ \alpha, \beta_m, \beta_a$
while $epoch$ **do**
  for $mini\text{-}batch\ in$ **Dataset do**
    $G_i^m \leftarrow L_{main}^i; G_i^a \leftarrow L_{aux}^i\ \forall i; i = [1, N] \wedge i \in \mathbb{N}$
    $\mu_m \leftarrow \frac{1}{N}\sum_{i=1}^{N} G_i^m, \sigma_m^2 \leftarrow \frac{1}{N}\sum_{i=1}^{N}(G_i^m - \mu_m)^2$
    $\mu_a \leftarrow \frac{1}{N}\sum_{i=1}^{N} G_i^a, \sigma_a^2 \leftarrow \frac{1}{N}\sum_{i=1}^{N}(G_i^a - \mu_a)^2$
    $calculate\ \zeta_m, \zeta_a, \tau(m, a)\ from\ \mu_m, \sigma_m, \mu_a, \sigma_a$
    $\rho(m, a) \leftarrow (\zeta_m/\zeta_a) \times \tau(m, a)$
    $\Delta\theta \leftarrow \beta_m\mu_m + \rho(m, a)\beta_a\mu_a$
    $\theta \leftarrow \theta + \alpha\Delta\theta$
    $\phi_{main} \leftarrow \phi_{main} + \alpha\beta_m\mu_m$
    $\phi_{aux} \leftarrow \phi_{aux} + \alpha\beta_a\mu_a$
  end for
end while

3. EXPERIMENTAL RESULTS

KITTI VO dataset [25] is used to learn pose estimation. This dataset contains 11 sequences; seq. 0-3, 6, 8-9 have been used for training while 4, 5, 7 and 10 have been used for validation. However, KITTI VO dataset does not include OF images for these sequences. We have trained FlowNetS [24] architecture separately using KITTI flow dataset and used the trained FlowNet to generate OF ground truth images for VO dataset. OF templates are trained using root mean squared logarithmic error (RMSLE) [10] to learn the OF subspace effectively (Figure 4), where 4(a) and 4(b) is consecutive image pair, 4(c) ground truth, 4(d) predicted OF. For model training, We have used (320×128) images with batch size 16. After hyperparameter tuning, we have found the best fitted model with learning rate 0.0005, $\beta_t = 1$, $\beta_r = 10$ and $\beta_a = 0.1$, gradient clipping has been applied to prevent overfitting. TensorFlow and Keras DL framework have been used for all experiments with machine specifications: Intel Core i7-9750H CPU@2.60GHz (12 CPUs), 16 GB RAM and NVIDIA GeForce GTX 1660Ti GPU.



(a) (b) (c) (d)

Figure 4. Optical flow subspace estimation (a),(b) image pair (c), ground truth, and (d) prediction

Figure 5 demonstrates (from left to right): i) simple NN method (without ATG), ii) vanilla ATG method, iii) PF method, iv) PRF method, v) Library for Visual Odometry 2 - Monocular (LIBVISO2-M) method, vi) Oriented FAST and rotated BRIEF-simultaneous localization and mapping (ORB-SLAM2) method. Here, v) [6] and vi) [26] are popular geometric methods commonly used for solving VO problem.They are given to compare PF and PRF methods' performance with respect to existing geometric methods. From i) to iv), it is evident that both PF and PRF method boosts the performance of simple NN as well as vanilla ATG method with a good margin (Figure 5, Table 2). Thus the claim regarding parameter reduction with our method is justified. It also outperforms other state-of-the-art (SOTA) ATL methods i.e the cosine similarity-based approach [14], and projecting conflicting gradients (PCGrad) [13] for MTL (Table 3). For fair comparison, we also modified PCGrad by only keeping the gradients directing the common normal for auxiliary task while keeping main task gradients unchanged. We referred this method as PCGrad ATL; which also cannot outperform ours. These results prove PRF method's effectiveness and superiority as an ATG method.

Table 3 demonstrates the effectiveness of our method for ATG and the comparison shows that our method outperforms the other SOTA ATL methods. Since the proposed method is applied to the complex problem of VO, comparison is also shown with some classic VO methods (Table 4) as well as SOTA VO methods (Table 5). The superiority of geometric and unsupervised learning-based approaches in the field of VO is undeniable. We acknowledge that our results are good but clearly do not beat the SOTA VO methods. However, the goal of this paper is not to outperform all the SOTA methods of VO, rather to show that using the proposed ATG method a complex problem like monocular VO can be solved with a remarkably smaller network while maintaining a relatively competitive results (Table 5) in supervised regime. Our inference network for pose estimation has 9,438,630 number of parameters which takes about 0.031 sec in average for each prediction. It has beaten most of the existing supervised methods in memory requiring at least 5-20 times less parameters. While most of the successful supervised methods highly exploits the temporal relation between frames by utilizing long sequences i.e 3-11 along with LSTM layers feeding high resolution images i.e 1280x384 (Table 5), we use dense layers for pose estimation and only one pair of images (320x128), which is the key reason of our faster inference. To our best knowledge, no supervised learning method can achieve this level of accuracy with such small parameter space.

It is evident that our method falls behind to some extent in case of the translation error. This is because like other supervised methods, it tries to learn absolute scale automatically from training images but at such a low parameter space absolute scale is not being learn very well. Some geometric methods takes advantage of loop closure and 7-DOF alignment with ground truth for scale correction. Future research can be done utilizing additional auxiliary task like depth estimation for better learning of absolute scale.

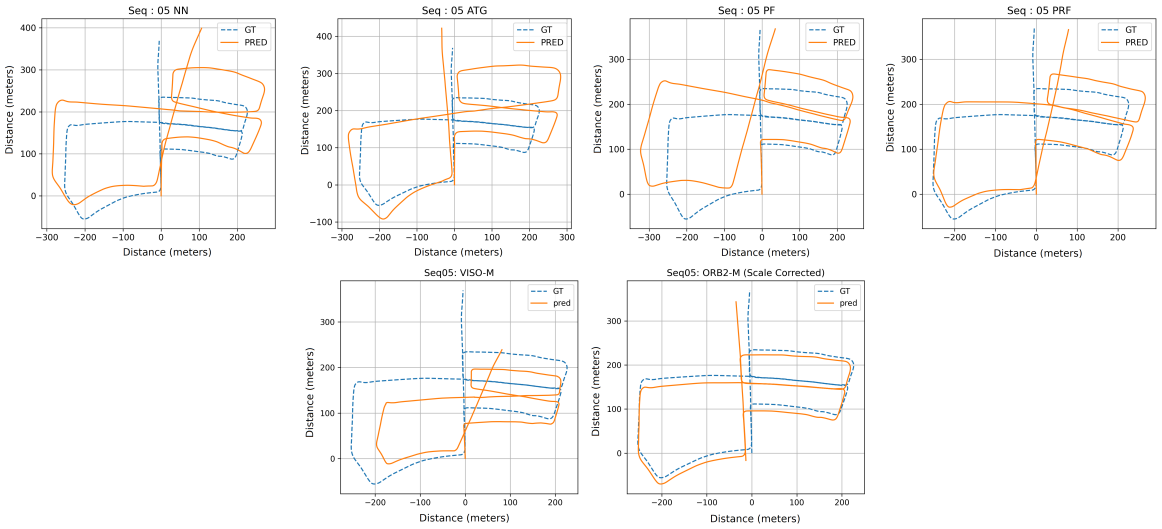


Figure 5. Comparative results of different approaches for sequence 05. Horizontal and vertical axes represent corresponding distances in the map.

Table 3. Comparison with existing methods of ATL methods

Seq.	Cosine Similarity [14]		PCGrad MTL [13]		PCGrad ATL		PF [Ours]		PRF [Ours]	
	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}
04	16.49	1.64	17.46	4.70	18.54	0.84	11.76	0.91	15.32	1.13
05	10.86	3.36	14.04	5.21	9.99	2.74	9.47	1.89	6.90	1.50
07	6.35	3.41	16.19	10.81	7.95	2.54	8.02	3.72	6.91	2.52
10	16.82	3.13	23.35	6.94	16.90	3.24	13.54	4.21	11.60	3.47

Table 4. Comparison of our results with some classic(p) methods in the field of VO

Seq.	LIBVISO2-M (G)		ORB-SLAM(G)		DeepVO (S)[17]		UnDeep VO(U)[22]		SfmLearner (U)[23]		PRF (S) [Ours]	
	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}
04	4.69	4.49	1.41	0.14	7.19	6.97	5.49	2.13	4.49	5.24	15.32	1.13
05	19.22	17.58	13.21	0.22	2.62	3.61	3.40	1.50	18.67	4.10	6.90	1.50
07	23.61	19.11	10.96	0.37	3.91	4.60	3.15	2.48	21.33	6.65	6.91	2.52
10	41.56	32.99	3.71	0.30	8.11	8.83	10.63	4.65	4.49	14.33	11.60	3.47

G: Geometric, S: Supervised, U: Unsupervised

Table 5. Comparison with SOTA supervised visual odometry methods

Arch.	DeepVO [17]		ESP-VO [19]		GFS-VO-RNN[20]		GFS-VO [20]		Beyond tracking[21]		PRF [Ours]	
Result	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}	t_{rel}	r_{rel}
Seq.04	7.19	6.97	6.33	6.08	5.95	2.36	**2.91**	1.30	2.96	1.76	15.32	**1.13**
Seq.05	2.62	3.61	3.35	4.93	5.85	2.55	3.27	1.62	**2.59**	**1.25**	6.90	1.50
Seq.07	3.91	4.60	3.52	5.02	5.88	2.64	3.37	2.25	**3.07**	**1.76**	6.91	2.52
Seq.10	8.11	8.83	9.77	10.2	7.44	3.19	6.32	2.33	**3.94**	**1.72**	11.60	3.47
Param.	463 M		463 M		**80 M		**47 M		**47 M		**9 M**	
Res.	1280x384		1280x384		1280x384		1280x384		1280x384		**320x128**	
Sq.len	Arbitary		Arbitary		7		7		11		**1**	

Param. : ** minimum possible parameters estimated based on available information, actual architecture may have higher number of parameters.

4. CONCLUSION

The PF method solved the issue of instability in learning. But it is prone to diminished probability which was solved by the PRF method. The PRF method additionally nullified common sources of variances successfully. Future works can include applying the proposed method for other fields and increasing the computational efficiency of the proposed methods.

REFERENCES

[1] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint*, Jun. 2017, doi: 10.48550/arXiv.1706.05098.

[2] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021, doi: 10.1109/TKDE.2021.3070203.

[3] S. Poddar, R. Kottath, and V. Karar, "Evolution of visual odometry techniques," *arXiv preprint*, 2018, doi: 10.48550/arXiv.1804.11142.

[4] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: large-scale direct monocular SLAM," in *European Conference on Computer Vision*, vol. 8690, 2014, pp. 834–849, doi: 10.1007/978-3-319-10605-2˙54.

[5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015, doi: 10.1109/TRO.2015.2463671.

[6] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium(IVS), Proceedings*, 2011, pp. 963–968, doi: 10.1109/IVS.2011.5940405.

[7] V. Guizilini and F. Ramos, "Visual odometry learning for unmanned aerial vehicles," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2011, pp. 6213–6220, doi: 10.1109/ICRA.2011.5979706.

[8] T. A. Ciarfuglia, G. Costante, P. Valigi, and E. Ricci, "Evaluation of non-geometric methods for visual odometry," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1717–1730, 2014, doi: 10.1016/j.robot.2014.08.001.

[9] V. Mohanty, S. Agrawal, S. Datta, A. Ghosh, V. D. Sharma, and D. Chakravarty, "DeepVO: a deep learning approach for monocular visual odometry," *arXiv preprint*, 2016, doi: 10.48550/arXiv.1611.06069.

[10] G. Costante and T. A. Ciarfuglia, "LS-VO: learning dense optical subspace for robust visual odometry estimation," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1735–1742, 2018, doi: 10.1109/LRA.2018.2803211.

[11] N. Yang, L. von Stumberg, R. Wang, and D. Cremers, "D3VO: deep depth, deep pose and deep uncertainty for monocular visual odometry," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 1278–1289, doi: 10.1109/CVPR42600.2020.00136.

[12] Z. Chen, V. Badrinarayanan, C. Y. Lee, and A. Rabinovich, "GradNorm: gradient normalization for adaptive loss balancing in deep multitask networks," in *35th International Conference on Machine Learning (ICML)*, 2018, vol. 2, pp. 794–803.

[13] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, "Gradient surgery for multi-task learning," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020, vol. 2020-Decem.

[14] Y. Du, W. M. Czarnecki, S. M. Jayakumar, M. Farajtabar, R. Pascanu, and B. Lakshminarayanan, "Adapting auxiliary losses using gradient similarity," *arXiv preprint*, Dec. 2018, doi: 10.48550/arXiv.1812.02224.

[15] G. Costante, M. Mancini, P. Valigi, and T. A. Ciarfuglia, "Exploring representation learning with CNNs for frame-to-frame ego-motion estimation," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 18–25, Jan. 2016, doi: 10.1109/LRA.2015.2505717.

[16] K. Konda and R. Memisevic, "Learning visual odometry with a convolutional network," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, 2015, vol. 1, pp. 486–490, doi: 10.5220/0005299304860490.

[17] S. Wang, R. Clark, H. Wen, and N. Trigoni, "DeepVO: towards end-to-end visual odometry with deep recurrent convolutional neural networks," in *Proceedings - IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2043–2050, doi: 10.1109/ICRA.2017.7989236.

[18] N. Yang, R. Wang, J. Stückler, and D. Cremers, "Deep virtual stereo odometry: leveraging deep depth prediction for monocular direct sparse odometry," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 817–833.

[19] S. Wang, R. Clark, H. Wen, and N. Trigoni, "End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks," *International Journal of Robotics Research*, vol. 37, no. 4–5, pp. 513–542, 2018, doi: 10.1177/0278364917734298.

[20] F. Xue, Q. Wang, X. Wang, W. Dong, J. Wang, and H. Zha, "Guided feature selection for deep visual odometry," in *Asian Conference on Computer Vision (ACCV)*, vol. 11366 LNCS, 2019, pp. 293–308.

[21] F. Xue, X. Wang, S. Li, Q. Wang, J. Wang, and H. Zha, "Beyond tracking: selecting memory and refining poses for deep visual odometry," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 8567–8575, doi: 10.1109/CVPR.2019.00877.

[22] R. Li, S. Wang, Z. Long, and D. Gu, "UnDeepVO: monocular visual odometry through unsupervised deep learning," in *Proceedings - IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 7286–7291, doi: 10.1109/ICRA.2018.8461251.
[23] L. Zhang, G. Li, and T. H. Li, "Temporal-aware SfM-learner: unsupervised learning monocular depth and motion from stereo video clips," in *Proceedings - 3rd International Conference on Multimedia Information Processing and Retrieval, MIPR 2020*, 2020, pp. 253–258, doi: 10.1109/MIPR49039.2020.00059.
[24] P. Fischer et al., "FlowNet: learning optical flow with convolutional networks," in Proceedings of the *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2758–2766.
[25] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013, doi: 10.1177/0278364913491297.
[26] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017, doi: 10.1109/TRO.2017.2705103.

BIOGRAPHIES OF AUTHORS

Irfan Mohammad Al Hasib is currently working as an Artificial Intelligence Engineer in Japan. He completed his B.Sc. degree from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh in Mechanical Engineering in 2017. His research area is centered to the field of computer vision, deep learning, embedded system and robotics. He is always passionate about designing intelligent system in the domain of computer vision and robotics. Further info on his homepage: https://irfanhasib0.github.io/. He can be contacted at email: irfanhasib.me@gmail.com.

Sumaiya Saima Sultana is currently working as a Machine Learning Researcher in Japan. She obtained her B.Sc in Mechanical Engineering from Bangladesh University of Engineering and Technology (BUET, Bangladesh) in 2018. Her research is focused on ML quality assurance, ML robustness analysis, deep learning,computer vision and robotics. Further info on her homepage: http://sumaiyasaima05.github.io/. She can be contacted at email: sumaiyasaima.sultana@gmail.com.

Imrad Zulkar Nyeen is currently working as an Artificial Intelligence Research Engineer in Japan. He obtained his B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET, Bangladesh) in 2018. His research interests include image processing, deep learning, machine learning applications in biomedical engineering and quality assurance of AI systems. He can be contacted at email: zulkareee13@gmail.com.

Muhammad Abdus Sabur is currently working as an Artificial Intelligence Engineer in Japan who completed his B.Sc. degree from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh in Electrical and Electronic Engineering in 2018. His research area is computer vision applications, deep learning, model deployment on edge device, combinatorial optimization and signal processing. He can be contacted at email: zihad146@gmail.com.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 137~145
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp137-145

Vehicle make and model recognition using mixed sample data augmentation techniques

Talha Anwar[1], Seemab Zakir[2]
[1]Center of Chiropractic Research, New Zealand College of Chiropractic, Auckland 1149, New Zealand
[2]Department of Engineering Technology, Foundation University, Rawalpindi, Pakistan

Article Info

Article history:

Received Sep 28, 2021
Revised Jul 7, 2022
Accepted Aug 5, 2022

Keywords:

Deep learning
Mixed data augmentation
Vehicle identification system

ABSTRACT

Vehicle identification based on make and model is an integral part of an intelligent transport system that helps traffic monitoring and crime control. Much research has been performed in this regard, but most of them used manual feature extraction or ensemble convolution neural networks (CNNs) that result in increased execution time during inference. This paper compared three deep learning models and utilized different augmentation techniques to achieve state-of-the-art performance without ensembling or fusing the models. Experimentations are made without any augmentation, with standard augmentation, and by mixed sample data augmentation techniques. Gradient accumulation and stochastic weighted averaging with mixed precision are used to have a large batch size that helped to reduce training time. The dataset comprised 48 vehicles' models running on the road of Pakistan. The highest accuracy and F1 score of 97% and 95% using the FMix augmentation technique with EfficientNetV2-S architecture gave the confidence that the proposed solution can be implemented in production.

Corresponding Author:

Talha Anwar
Center of Chiropractic Research, New Zealand College of Chiropractic
Auckland 1149, New Zealand
Email: chtalhaanwar@gmail.com

1. INTRODUCTION

Vehicle identification system (VIS), an integral component of the intelligent transport system (ITS), brings ease to the traffic management system and helps against criminal activities. VIS is widely used in road violation detection, traffic congestion alarm, and unmanned driving. Millions of vehicles are on the road in big cities, making it challenging to track a particular vehicle. The vehicles' number plate is mostly used to track them [1], but number plates can be changed easily, leading to false identification. VIS also helps automate tax collection at toll plazas based on vehicle type.

With the advent of artificial intelligence (AI), deep learning has been widely used in transportation [2] Some recent studies used traditional imaging techniques such as haar-like features with AdaBoost classifier [3] and pattern descriptors with support vector classifier [4]. The pattern descriptors study used local binary patterns, median binary patterns, directional gradient patterns, and local arc patterns as features. Kiran *et al.* also studied different colour spaces such as red, green and blue (RGB), green (Y), blue (Cb), red (Cr) (YcbCr) and hue, saturation, value (HSV) for descriptor extraction [4] haar-like features-based study first removed shadows using HSV colour space to reduce the chances of false detection. Different single feature methods, such as colour moment, local binary pattern (LBP) features, Hu moment features, angle features, and circularity are also used. Using Adaboost 85.8% accuracy is achieved [3]. Qiu *et al.* [5] compared the performance of haar features along with convolution neural network (CNN). Using haar-like

features, 86.72% and 91.86% precision and recall are achieved, which increased by 5.63% and 0.2% with CNN [5]. Gholamalinejad and Khosravi proposed a novel CNN architecture composed of CNN layers with squeeze-and-excitation (SE) modules. Instead of using classic max pooling or average pooling, they used haar wavelet as a pooling layer [6]. The data is composed of 5 classes, including bus, heavy truck, medium truck and pickup. They achieved an accuracy of 95.1% [6]. Ajitha *et al.* proposed a shallow CNN model with traditional augmentation techniques such as flip, rotation, shear, crop and zoom, resulting in an accuracy of 92.3% [7]. Mansor *et al*. [8] achieved an accuracy of 95% with 4 class classification problems. Their work is based on emergency vehicle type classification and had images of fire trucks, police cars, ambulances and standard cars [8]. Hassan *et al*. compared different classifiers with cyclic learning rate and used the MixUp image augmentation technique to achieve an accuracy of 93.96% through ensembling homogeneous models of DenseNet201 [9]. Though the CNN-based model has gained much attention in recent years, manual feature-based classification is still being studied recently. Chen detected multiple features from the vehicle, such as taillight features, shadow area features and other descriptors. Radial basis function (RBF) artificial neural network is further used for classification and achieved 97% accuracy [10]. Another manual feature-based study used histogram-oriented gradients (HOG) and ant colony optimization (ACO) to classify vehicles and achieved an accuracy of 90% [11].

All the existing studies either deal with a few vehicle models, manual features extraction or used ensemble models in which multiple models are tested during inference resulting in increased prediction time. As the VIS is implemented in real-time, it needs to be robust. Keeping in view the limitation, we proposed a single network-based approach that yields the state of the art performance. Three different models and five augmentations techniques are compared. All the experiments are seeded for the purpose of reproducibility. The main contributions of this paper are,

- Different deep learning architectures are compared without using any augmentation technique, with commonly used and mixed sample data augmentation techniques (MSDA).
- Ensemble and fusion of different models increase the inference time, so the approach used a single model that performed better than the existing ensembled models.
- The proposed approach achieved state-of-the-art performance with 97% and 95% accuracy and F1 score, respectively.

The paper is organized: The introduction, motivation, and literature review on vehicle classification are presented in section 1. Section 2 describes the methodology in detail. Section 3 deals with results and discussion. The conclusion is made in section 4. The implementation is publicly available at GitHub [12].

2. METHOD

2.1. Dataset

We used images of common cars running on the road of Pakistan [13]. There are 3,103 and 752 training and test images divided into 48 car models/classes. Figure 1 shows the sample image. Table 1 shows the vehicle name and the number of images available for training for each vehicle.

2.2. Transformation

Transformation is a technique to produce variation in the data. It helps to generalize prediction on test data and avoid over-fitting the model. Albumentation [14] library is used for this purpose. Following the main standard Augmentation used for applied transformations:

- Resize: all images are resized to 256×256
- Center crop: crop all images are centre cropped to 224×224
- Horizontal Flip: fifty per cent of images are horizontally flipped
- Vertical Flip: fifty per cent of images are flipped vertically
- Shift scale rotate: fifty per cent of images are randomly shifted, rotated, and scaled in height and width.
- CLAHE: contrast limited adaptive histogram equalization (CLAHE) is a modified form of adaptive histogram equalization. In histogram equalization, the intensity range of the image is stretched between 0 and 255 to improve the contrast of the image. However, this led to either too dark or too bright picture. Adaptive histogram handled this issue by dividing the image into small patches and applied histogram equalization on each patch. This sometimes led to over-amplification of contrast if the image has noise. CLAHE performed bi-linear interpolation on the edges of patches and reduced this contrast amplification by removing the artificial boundaries.
- Cutout: cutout is one of the ways to handle over-fitting. In this technique, black boxes are introduced in images, making the image classification hard, and reduced the chances of over-fitting.
- Normalization: normalization led to fast convergence and speeds up the training process.

Figure 1. Sample vehicles image from each class label, the number on each image corresponds to the vehicle ID in Table 1

Table 1. Vehicle models and the number of images for that models. ID column is related to Figure 1. No. shows number of training examples for that model

ID	Vehicle model	No
1	Daiatsu Core	80
2	Daiatsu Hijet	44
3	Daiatsu Mira	81
4	FAW V2	29
5	FAW XPV	26
6	Honda BRV	27
7	Honda city 1994	32
8	Honda city 2000	69
9	Honda City aspire	105
10	Honda civic 1994	16
11	Honda civic 2005	34
12	Honda civic 2007	74
13	Honda civic 2015	31
14	Honda civic 2018	82
15	Honda Grace	21
16	Honda Vezell	38
17	KIA Sportage	25
18	Suzuki alto 2007	132
19	Suzuki alto 2019	56
20	Suzuki alto japan 2010	27
21	Suzuki carry	13
22	Suzuki cultus 2018	269
23	Suzuki cultus 2019	108
24	Suzuki Every	20
25	Suzuki highroof	63
26	Suzuki kyber	52
27	Suzuki liana	33
28	Suzuki margala	16
29	Suzuki Mehran	195
30	Suzuki swift	118
31	Suzuki wagonR 2015	112
32	Toyota hiace 2000	23
33	Toyota Aqua	77
34	Toyota axio	20
35	Toyota corolla 2000	39
36	Toyota corolla 2007	82
37	Toyota corolla 2011	127
38	Toyota corolla 2016	270
39	Toyota fortuner	43
40	Toyota Hiace 2012	72
41	Toyota Landcruser	17
42	Toyota Passo	61
43	Toyota pirus	23
44	Toyota Prado	21
45	Toyota premio	18
46	Toyota Vigo	53
47	Toyota Vitz	81
48	Toyota Vitz 2010	48

2.3. Mixed sample data augmentation

Large neural networks are notorious for memorizing data instead of learning it even in strong regularization and fail during inference. Though standard data augmentation helped in generalization, this technique is data-dependent and required domain knowledge. Anwar and Zakir [15] studied that standard augmentation sometimes led to poor results. They explored different image augmentation techniques on electrocardiogram (ECG) graphs and found that the best results are obtained without applying any augmentation. CNN focused on the discriminative part of the image instead of the whole image leading to poor generalization. Regional dropout techniques such as the CutOut helped the CNN to view the bigger image perspective, but this reduced the proportion of informative pixels of training data [16]. Mixed Sample data augmentation (MSDA) techniques are introduced to overcome standard augmentation and generalization issues. MSDA mixed different distributions of data to produce new data from the same distribution of existing data. It is categorized into two policies, interpolation and masking. MixUp is an example of interpolation, whereas CutMix and FMix are an example of masking MSDA.

2.3.1. Mixup

MixUp mixed two images from different classes and linearly interpolated them to produce a new image. It not only interpolated the input images' features but also interpolated the corresponding target [17]. The working principle of MixUp is shown in (1) and (2),

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \quad (1)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \quad (2)$$

x_i and x_j are raw images in (1) and y_i and y_j are the one-hot encoded labels in (2). λ drawn from β distribution is used to mix two random images. MixUp increased the capability of deep learning architectures to learn from corrupted labels and improved the generalization. Linear interpolation of input images reduced the memorization by large deep learning models [18].

2.3.2. CutMix

Cutout and MixUp inspired CutMix paper. It claimed to resolve the issues in MixUp. Though MixUp improved classification performance, the resulting sample is unnatural. CutMix replaced an image patch with a patch of another random picture from the training data [16]. It is like a cutout where a patch is replaced with zeros and MixUp where two images are mixed.

$$\tilde{x} = M x_i + (1 - M) x_j \quad (3)$$

Patch mixing in training images is shown in (3). M is a binary mask indicating where the dropout rectangular region should be placed. Then this rectangular dropout region is replaced by a patch of another image. Mixing of one-hot encoded labels is the same as in the MixUp technique. CutMix focused on the less discriminative part of the object, whereas Mixup focused on the entire image but produced unnatural artefacts.

2.3.3. FMix

CutMix reduced overfitting by increasing the observable data points without changing the data distribution. However, CutMix used square patches, which is a limitation and leads to distortion. FMix claimed to resolve the issue in CutMix by using binary masks obtained by applying a threshold to low-frequency images from the Fourier space. The authors first sampled low-frequency grayscaled masks from Fourier space and then converted them to binary masks using a threshold. Once a binary mask is obtained, two images from different classes are overlaid together, such as 0 pixels of binary mask corresponded to one image and pixels with 1 value of binary mask is related to another image from a different class. FMix, unlike CutMix, proposed patches of different shapes which maximize the number of possible masks [19].

Overall, when data is limited and learning from individual examples is easier, MixUp is a good candidate, and FMix is a better choice when data is abundant. In Figure 2, MixUp shows that two images are mixed together in an overlay fashion. CutMix shows that a square patch of another image replaces a square patch. FMix shows that another image from the training data replaced a randomly shaped patch of an image.

2.4. Deep learning architecture

Deep learning is a subset of artificial intelligence that takes the complex raw data as input, automatically extracts valuable features, and performs task-relevant work such as classification or regression.

In image classification, deep learning boomed in 2014 after VGGNet came out. Though before VGG, AlexNet was there, VGG16 outperformed it by 10%. At that time, it was believed that increasing the layer increased the performance of the model, until in December 2015, ResNet paper was released and proved that adding layers helped to some extent and started decreasing the performance beyond that [20]. To date, ResNet or ResNet variants are one of the most used architecture; therefore, we decided to use ResNet as our baseline.

Figure 2. Mixed sample data augmented images of two cars

2.4.1. ResNet

Ideally, a deeper neural network is preferable as it yields better results. Nevertheless, this comes with the cost of vanishing gradient and degradation. By increasing the depth of the neural network, the gradients became very small during back-propagation and reached zero; this phenomenon is known as vanishing gradient. Though this problem can be resolved using the rectified linear units (ReLU) activation function, skip connection also played a role. Skip connection back-propagates the gradient of larger magnitude by skipping some layers in between.

ResNet paper explained that further deepening neural network led to a significant error rate characterized by degradation. Adding layers saturated the model, and the error rate started increasing. It is believed that if a shallow network is working fine, the additional deep layers should work the same though it did not happen, and deep networks start performing poorly. So, an identity function is added from a shallow layer to a deeper layer, and the model started learning that identity function. In ResNet, this identity function ensured that the deep network output should be identical to the shallow network. ResNet paper named this identity function as skip connections that skip some layers and pass information directly to other layers by an identity function. In the worst case, the performance of a deeper network will not be worse than a shallow network, and in the best scenario, it can be better than the shallow network [20]. Multiple ResNet variants are described by network size and the number of layers skipped by the skip connections. We used ResNet-50 as it is neither tiny to underfit nor very large to overfit.

2.4.2. DenseNet

DenseNet was proposed in 2018 by Huang *et al*. [21]. Based on the observation, if there is a shorter connection between input and output layers, the model can be deeper, more accurate, and more efficient to train. DenseNet is based on dense blocks and transition layers. In dense blocks, each coming layer received collective information from all previous layers both directly and indirectly. Similarly, in back-propagation, the error signal collectively flowed to all layers. For each layer, the feature maps of all previous layers are considered output, and the output of that layer is considered as input for all subsequent layers. For the sake of downsampling to reduce network size, a transition layer between two dense blocks is used. This layer is composed of a 1×1 convolution filter preceded and followed by batch normalization and an average pooling layer. We used DenseNet 121 in this study.

2.4.3. EfficientNetV2

Most of the deep learning architecture either scaled the depth such as ResNet by increasing the number of layers or width by adding more neurons/filters in each layer, for example, wide ResNet [22]. Wider networks learn more detailed features and are easier to train because they are usually shallower However, shallower and wider networks have an issue in learning high-level features. Some networks used high-resolution images such as InceptionV3 which used 299×299 image size [23]. Scaling a specific dimension such as depth, width, and resolution increase accuracy up to a limit. EfficientNet in 2019 claimed that its depth, width and resolution should be scaled proportionally to make a deeper network more effective. So the authors proposed a compound scaling method to scale width, depth and resolution proportionally [24].

EfficientNetV2 in June 2021 is one of the latest proposed models and is known for faster training speed [25]. This model is based on training awareness neural architecture search (NAS) and progressive scaling. It is observed that small image sizes require less regularization as compared to large image sizes. So the authors started with small image size and increased the size progressively. They used EfficientNet as their backbone architecture and applied the NAS strategy, though the authors removed unnecessary search options to reduce the search space. This paper used a small kernel size of 3×3 and added more layers to compensate for the reduced receptive field. Other tweaks are applied to reduce the memory access overhead in EfficientNet, such as removing the last stride layer. In our study, EfficientNetV2-S is used.

2.5. Explainability of MSDA techniques

To understand the impact of MSDA techniques, we used gradient-weighted class activation mapping (Grad-CAM) that explained which area of an image is focused by a network to decide the label class. Grad-CAM produced a localization heatmap of the target by utilizing its gradient against the last convolution layers and highlighted the essential regions of the image [26]. To generate Grad-CAM PyTorch library for CAM methods is used [27].

2.6. Additional information

Fifty epochs are trained with a learning rate and batch size of 0.001 and 48, respectively. AdamW optimizer is used instead of Adam as it provides better results [15]. Pytorch Lightning framework is used for implementation. Accuracy, macro F1 score, precision and recall are used for evaluation. Mixed precision, gradient accumulation, and stochastic weight averaging (SWA) techniques are used to speed up the training time. Gradient accumulation is a technique to train the model with larger batch sizes by updating weights after some batches instead of every batch. SWA helps to generalize the model, whereas Mixed precision reduces training time up to 8x [28] by allowing a large batch size.

3. RESULTS AND DISCUSSION

This paper deals with the identification of commonly used vehicles in Pakistan. Table 2 shows the performance of different augmentation techniques with three deep learning architectures. Without using any augmentation technique, an F1 score of 88%,91%, and 90% is achieved using ResNet-50, DenseNet121 and EfficientNetV2-S, respectively. When standard augmentations are applied, the F1 score increased in all three models, which shows the impact of data augmentation. With MixUp augmentation techniques in which two images are mixed together in an overlay fashion, there is not much difference in the F1 score of different deep learning models compared with standard augmentations. When CutMix is applied, there is 1% increment in accuracy obtained using EfficientNet and ResNet. FMix augmentation technique achieved the highest accuracy and F1 score in all deep learning models. EfficientNetV2 with FMix augmented input resulted in accuracy and F1 score of 97% and 95%, respectively. With EfficientNetV2 this is a 2% increment in F1 score compared to MixUp and CutMix augmentation techniques. Without augmentation, the macro F1 score is 90% which increased by 5% with FMix augmentation technique. These MSDA augmentation techniques are applied without standard augmentation to study the impact of MSDA augmentations alone. Figure 3 shows validation loss using five different augmentation techniques. The lowest validation loss is achieved using FMix augmentation technique when EfficientNetV2-S model is used. EfficientNetV2-S also showed the second-lowest curve with the CutMix MSDA technique. CutMix and MixUp produced similar results in standard augmentation, but FMix outperformed them in all three deep learning architectures.

Figure 4 shows the heatmap generated by the Grad-CAM technique. MixUp techniques paid attention to most parts of the car's front, but its focus is diverged. On the other hand, CutMix focused on the right front headlight, but its span of coverage is less. FMix covered both aspects, its heatmap is more focused and spread over the front area. It helped the model visualize and focus broader region while making a decision and providing better results.

The existing studies are either based on manual features extraction [3] or multiple ensemble models [9] resulted in reduced performance during inference. The proposed solution is robust during inference but has some limitations during training. The more the augmentation, the more time a model needs to train itself because an image undergoes a series of transformations before feeding to the neural network. We observed that MSDA augmentation takes time to do the mathematical calculation of image mixing. However, no augmentations are applied during test time, making the model robust during the inference.

The limitation of standard augmented CNN or features-based classifiers is adversarial image attacks. Manipulating certain car parts can make CNN fool, and it would not predict the vehicle. On the other hand, MSDA techniques heavily altered the image by placing other pictures on it; thus, there would be minimal chances of adversarial attacks. FMix resolved the issues of CutMix which is inspired by MixUp, so

theoretically, FMix should have better performance [19]. Practically this is proved as FMix augmentation got 1%, 2% and 2% accuracy improvement in EfficientNetV2-S, DenseNet121 and ResNet50 as compared to CutMix, respectively.

Table 2. Model performance using different augmentations techniques

	ResNet-50				DenseNet121				EfficientNet			
Techniques	F1	Prec	Rec	Acc	F1	Pre	Rec	Acc	F1	Prec	Rec	Acc
None	88%	90%	87%	92%	91%	94%	91%	94%	90%	92%	88%	94%
Standard	90%	91%	90%	93%	92%	93%	91%	94%	93%	95%	92%	95%
MixUp	90%	94%	89%	94%	91%	94%	90%	94%	93%	96%	92%	95%
CutMix	91%	94%	90%	95%	91%	94%	90%	95%	93%	96%	92%	96%
FMix	93%	94%	92%	95%	94%	95%	94%	97%	95%	96%	95%	97%

Prec: precision, Rec: recall, F1: f1 score, Acc: accuracy

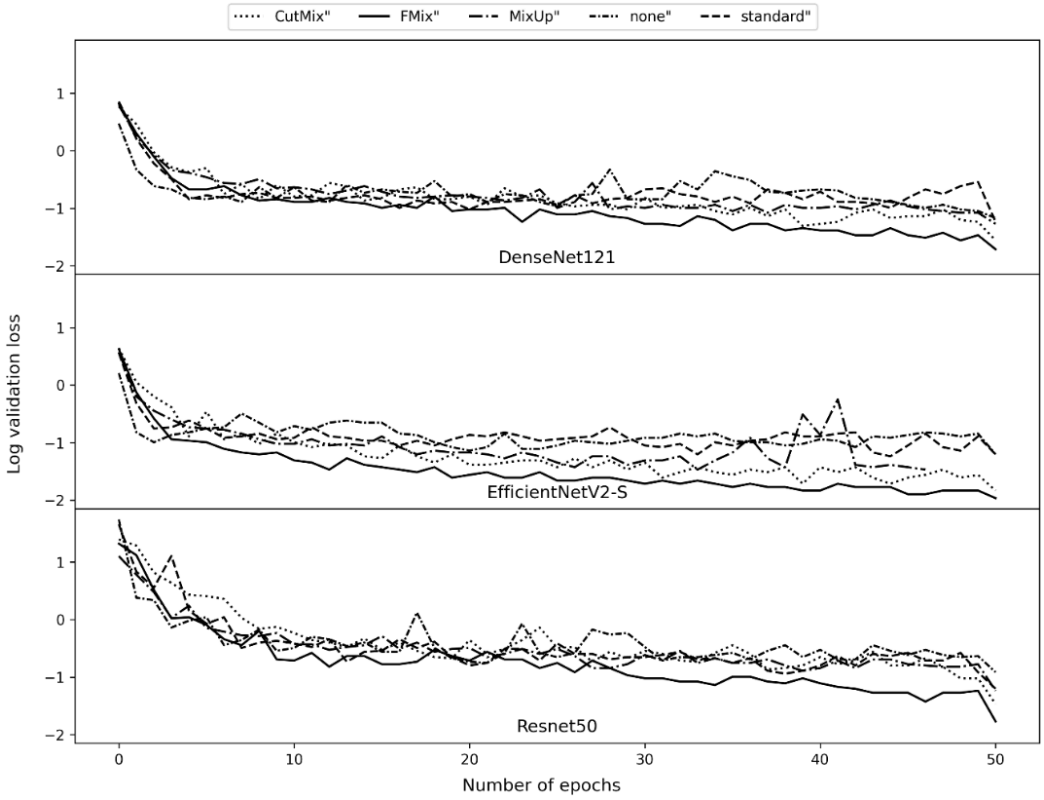


Figure 3. Validation loss using different architectures and augmentation techniques. Three different subplots with a common axis show three deep learning architectures. Five different patterns show five different augmentation methods

Figure 4. Grad-CAM heatmap for MSDA augmentation techniques

4. CONCLUSION

In this paper, different augmentation techniques are studied to achieve the state of art results. Unlike other studies that used manual feature extraction such as edge detection or haar features, this study used end-to-end CNN to extract and classify features automatically. Ensemble models are not used because they are not feasible for deployment because of time complexity and inference time limitations. Five augmentation scenarios are used, such as no augmentation, standard augmentation, and three mixed sample data augmentation techniques. Three deep learning algorithms such as ResNet, DenseNet and EfficientNet are used. All five augmentation techniques and three CNN architectures are compared. Mixed sample data augmentation techniques helped to achieve state-of-the-art performance using an EfficientNetV2-S model on a dataset comprised of 48 models of vehicles running on the roads of Pakistan. Further, the heatmap of MSDA techniques are compared to understand the learning of deep learning model. FMix image augmentation with EfficientNetV2 resulted in the highest F1 score of 95%, which is 5% better if no augmentation is applied and 2% better if standard commonly used augmentation techniques are used.

REFERENCES

[1] P. N. Huu and C. V. Quoc, "Proposing WPOD-NET combining SVM system for detecting car number plate," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, p. 657, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp657-665.

[2] Y. Wang, D. Zhang, Y. Liu, B. Dai, and L. H. Lee, "Enhancing transportation systems via deep learning: a survey," *Transportation Research Part C: Emerging Technologies*, vol. 99, pp. 144–163, 2019, doi: 10.1016/j.trc.2018.12.004.

[3] L. Zhang, J. Wang, and Z. An, "Vehicle recognition algorithm based on Haar-like features and improved Adaboost classifier," *Journal of Ambient Intelligence and Humanized Computing*, 2021, doi: 10.1007/s12652-021-03332-4.

[4] V. Keerthi Kiran, S. Dash, and P. Parida, "Vehicle recognition using extensions of pattern descriptors," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1166, no. 1, p. 12046, doi: 10.1088/1757-899x/1166/1/012046.

[5] L. Qiu, D. Zhang, Y. Tian, and N. Al-Nabhan, "Deep learning-based algorithm for vehicle detection in intelligent transportation systems," *Journal of Supercomputing*, vol. 77, no. 10, pp. 11083–11098, 2021, doi: 10.1007/s11227-021-03712-9.

[6] H. Gholamalinejad and H. Khosravi, "Vehicle classification using a real-time convolutional structure based on DWT pooling layer and SE blocks," *Expert Systems with Applications*, vol. 183, 2021, doi: 10.1016/j.eswa.2021.115420.

[7] P. Ajitha, S. Jeyakumar, Y. N. Krishna K, and A. Sivasangari, "Vehicle model classification using deep learning," in *Proceedings of the 5th International Conference on Trends in Electronics and Informatics, ICOEI 2021*, 2021, pp. 1544–1548, doi: 10.1109/ICOEI51242.2021.9452842.

[8] M. A. Hakim Bin Che Mansor, N. A. Mohamad Kamal, M. H. Bin Baharom, and M. Adib Bin Zainol, "Emergency vehicle type classification using convolutional neural network," in *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings*, 2021, pp. 126–129, doi: 10.1109/I2CACIS52118.2021.9495899.

[9] A. Hassan, M. Ali, N. M. Durrani, and M. A. Tahir, "An empirical analysis of deep learning architectures for vehicle make and model recognition," *IEEE Access*, vol. 9, pp. 91487–91499, 2021, doi: 10.1109/ACCESS.2021.3090766.

[10] X. Chen, H. Chen, and H. Xu, "Vehicle detection based on multifeature extraction and recognition adopting RBF neural network on ADAS system," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8842297.

[11] R. S. El-Sayed and M. N. El-Sayed, "Classification of vehicles' types using histogram oriented gradients: comparative study and modification," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4, pp. 700–712, 2020, doi: 10.11591/ijai.v9.i4.pp700-712.

[12] T. Anwar, "Pak vehicle classification," *GitHub repository*. 2021.

[13] M. Ali, M. A. Tahir, and M. N. Durrani, "Vehicle images dataset for make and model recognition," *Data in Brief*, vol. 42, p. 108107, Jun. 2022, doi: 10.1016/j.dib.2022.108107.

[14] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information (Switzerland)*, vol. 11, no. 2, 2020, doi: 10.3390/info11020125.

[15] T. Anwar and S. Zakir, "Effect of image augmentation on ECG image classification using deep learning," in *2021 International Conference on Artificial Intelligence, ICAI 2021*, 2021, pp. 182–186, doi: 10.1109/ICAI52203.2021.9445258.

[16] S. Yun, D. Han, S. Chun, S. J. Oh, J. Choe, and Y. Yoo, "CutMix: regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-Octob, pp. 6022–6031, doi: 10.1109/ICCV.2019.00612.

[17] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: beyond empirical risk minimization," Apr. 2018, doi: 10.48550/arXiv.1710.09412.

[18] D. Liang, F. Yang, T. Zhang, and P. Yang, "Understanding mixup training methods," *IEEE Access*, vol. 6, pp. 58774–58783, 2018, doi: 10.1109/ACCESS.2018.2872698.

[19] E. Harris, A. Marcu, M. Painter, M. Niranjan, A. Prügel-Bennett, and J. Hare, "FMix: Enhancing Mixed Sample Data Augmentation," *arXiv preprint*, Feb. 2020.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/cvpr.2016.90.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 4700–4708, doi: 10.1109/cvpr.2017.243.

[22] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *British Machine Vision Conference 2016, BMVC 2016*, 2016, vol. 2016-Septe, pp. 87.1--87.12, doi: 10.5244/C.30.87.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[24] M. Tan and Q. V Le, "EfficientNet: rethinking model scaling for convolutional neural networks," *arXiv preprint*, May 2019, doi: 10.48550/arXiv.1905.11946.

[25] M. Tan and Q. V Le, "EfficientNetV2: smaller models and faster training," *arXiv preprint*, 2021.

[26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-Octob, pp. 618–626, doi: 10.1109/ICCV.2017.74.
[27] J. Gildenblat *et al.*, "PyTorch library for CAM methods," *GitHub*, 2021.
[28] S. Narang *et al.*, "Mixed precision training," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, pp. 1–12.

BIOGRAPHIES OF AUTHORS

Talha Anwar is an AI researcher having a Master's degree in Data Science from FAST, National University, Pakistan. He obtained Bachelor's Degree in Biomedical Engineering from Riphah International University in 2018. His research is in biomedical image analysis, biosignal analysis, particularly in the area of brain-computer interface. He has a special interest in social text analysis in the field of NLP. He is equally interested in machine learning and deep learning and has several publications in this domain. Talha is actively involved in research and working with Centre for Chiropractic Research, New Zealand College of Chiropractic, Auckland 1060, New Zealand. All of his research is available at github.com/talhaanwarch. He can be contacted at email: chtalhaanwar@gmail.com.

Seemab Zakir has Bachelor's and Masters's degrees in biomedical engineering from Riphah International University, Pakistan. She has experience in conducting labs on biomedical engineering subjects, particularly programming, machine learning, and instrumentation. She has also served as a biomedical engineer at Pak-Austria Fachhochschule: Institute of Applied Sciences. She was a lecturer at Foundation University School of Science and Technology, Pakistan. Currently, she is a Ph.D. scholar at Scuola Superiore Sant'Anna Pisa, Italy. Her areas of interest are biomedical instrumentation and artificial intelligence. She can be contacted at email: seemabzakir2@gmail.com.

IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 99, No. 1, Month 2099, pp. 1~1x
ISSN: 2252-8938, DOI: 10.11591/ijai.v99.i1.pp1-1x

Paper's title should be the fewest possible words that accurately describe the content of the paper (Center, Bold, 16pt)

Abdel-Rahman Hedar[1,2], Patricia Melin[3], Kennedy Okokpujie[4] (10 pt)
[1]Department of Computer Science, Faculty of Computers & Information, Assiut University, Assiut, Egypt (8 pt)
[2]Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah, Saudi Arabia
[3]Division of Graduate Studies, Tijuana Institute of Technology, Tijuana, Mexico
[4]Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ogun State, Nigeria

Article Info

Article history:

Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

Keywords:

First keyword
Second keyword
Third keyword
Fourth keyword
Fifth keyword

ABSTRACT (10 PT)

An abstract is often presented separate from the article, so it must be able to stand alone. A well-prepared abstract enables the reader to identify the basic content of a document quickly and accurately, to determine its relevance to their interests, and thus to decide whether to read the document in its entirety. The abstract should be informative and completely self-explanatory, provide a clear statement of the problem, the proposed approach or solution, and point out major findings and conclusions. **The Abstract should be 100 to 200 words in length.** References should be avoided, but if essential, then cite the author(s) and year(s). Standard nomenclature should be used, and non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself. No literature should be cited. The keyword list provides the opportunity to add 5 to 7 keywords, used by the indexing and abstracting services, in addition to those already present in the title (9 pt).

Corresponding Author:

Kennedy Okokpujie
Department of Electrical and Information Engineering, College of Engineering, Covenant University
Km. 10 Idiroko Road, Canaan Land, Ota, Ogun State, Nigeria
Email: kennedy.okokpujie@covenantuniversity.edu.nga

1. INTRODUCTION (10 PT)

The main text format consists of a flat left-right columns on A4 paper (quarto). The margin text from the left and top are 2.5 cm, right and bottom are 2 cm. The manuscript is written in Microsoft Word, single space, Time New Roman 10 pt, and maximum 12 pages for original research article, or maximum 16 pages for review/survey paper, which can be downloaded at the website: http://ijai.iaescore.com.

A title of article should be the fewest possible words that accurately describe the content of the paper. The title should be succinct and informative and no more than about 12 words in length. Do not use acronyms or abbreviations in your title and do not mention the method you used, unless your paper reports on the development of a new method. Titles are often used in information-retrieval systems. Avoid writing long formulas with subscripts in the title. Omit all waste words such as "*A study of ...*", "*Investigations of ...*", "*Implementation of ...*", "*Observations on ...*", "*Effect of.....*", "*Analysis of ...*", "Design of…", etc.

A concise and factual abstract is required. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided, but if essential, then cite the author(s) and year(s). Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself. Immediately after the abstract, provide a maximum of 7 keywords, using American spelling and avoiding general and plural terms and

multiple concepts (avoid, for example, 'and', 'of'). Be sparing with abbreviations: only abbreviations firmly established in the field may be eligible. These keywords will be used for indexing purposes.

Indexing and abstracting services depend on the accuracy of the title, extracting from it keywords useful in cross-referencing and computer searching. An improperly titled paper may never reach the audience for which it was intended, so be specific.

The Introduction section should provide: i) a clear background, ii) a clear statement of the problem, iii) the relevant literature on the subject, iv) the proposed approach or solution, and v) the new value of research which it is innovation (within 3-6 paragraphs). It should be understandable to colleagues from a broad range of scientific disciplines. Organization and citation of the bibliography are made in Institute of Electrical and Electronics Engineers (IEEE) style in sign [1], [2] and so on. The terms in foreign languages are written italic (*italic*). The text should be divided into sections, each with a separate heading and numbered consecutively [3]. The section or subsection headings should be typed on a separate line, e.g., 1. INTRODUCTION. A full article usually follows a standard structure: **1. Introduction, 2. The Comprehensive Theoretical Basis and/or the Proposed Method/Algorithm** *(optional)***, 3. Method, 4. Results and Discussion, and 5. Conclusion.** The structure is well-known as **IMRaD** style.

Literature review that has been done author used in the section "INTRODUCTION" to explain the difference of the manuscript with other papers, that it is innovative, it are used in the section "METHOD" to describe the step of research and used in the section "RESULTS AND DISCUSSION" to support the analysis of the results [2]. If the manuscript was written really have high originality, which proposed a new method or algorithm, the additional section after the "INTRODUCTION" section and before the "METHOD" section can be added to explain briefly the theory and/or the proposed method/algorithm [4].

2. METHOD (10 PT)

Explaining research chronological, including research design, research procedure (in the form of algorithms, Pseudocode or other), how to test and data acquisition [5]–[7]. The description of the course of research should be supported references, so the explanation can be accepted scientifically [2], [4]. Figures 1-2 and Table 1 are presented center, as shown below and cited in the manuscript [5], [8]–[13]. The settlement curves produced at SG1 has been illustrated in Figure 2(a) and SG2 has been illustrated Figure 2(b).

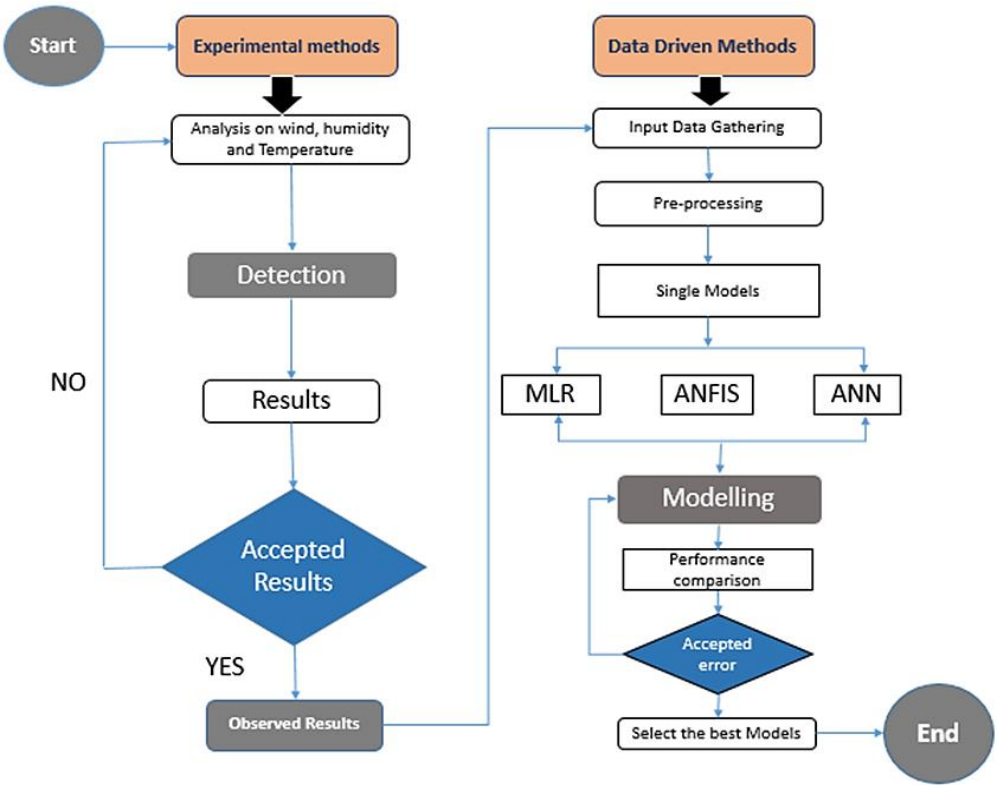
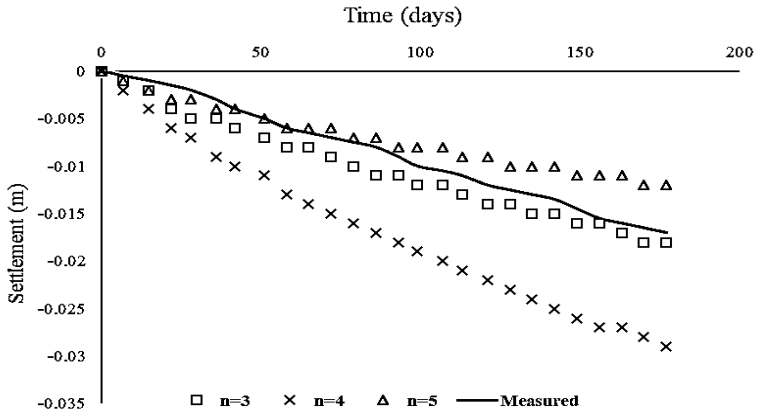
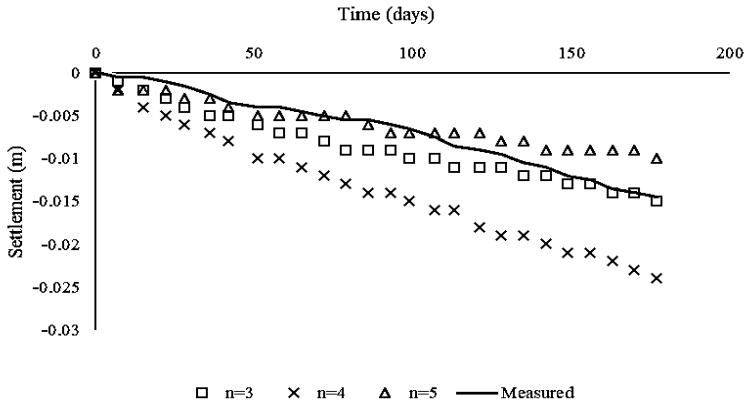


Figure 1. Shows the flowchart of the AI-based models and experimental methods applied



(a)



(b)

Figure 2. The relationship of soil settlement and time, (a) SG1 and (b) SG2

Table 1. The performance of ...

Variable	Speed (rpm)	Power (kW)
x	10	8.6
y	15	12.4
z	20	15.3

3. RESULTS AND DISCUSSION (10 PT)

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [14], [15]. The discussion can be made in several sub-sections.

3.1. Sub section 1

Equations should be placed at the center of the line and provided consecutively with equation numbers in parentheses flushed to the right margin, as in (1). The use of Microsoft Equation Editor or MathType is preferred.

$$E_v - E = \frac{h}{2.m} (k_x^2 + k_y^2) \quad (1)$$

All symbols that have been used in the equations should be defined in the following text.

3.2. Sub section 2

Proper citation of other works should be made to avoid plagiarism. When referring to a reference item, please use the reference number as in [16] or [17] for multiple references. The use of ”Ref [18]...”

should be employed for any reference citation at the beginning of sentence. For any reference with more than 3 or more authors, only the first author is to be written followed by *et al*. (e.g. in [19]). Examples of reference items of different categories shown in the References section. Each item in the references section should be typed using 9 pt font size [20]–[25].

3.2.1. Subsub section 1

yy

3.2.2. Subsub section 2

zz

4. CONCLUSION (10 PT)

Provide a statement that what is expected, as stated in the "INTRODUCTION" section can ultimately result in "RESULTS AND DISCUSSION" section, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

ACKNOWLEDGEMENTS (10 PT)

Author thanks ... . In most cases, sponsor and financial support acknowledgments.

REFERENCES (10 PT)

The main references are international journals and proceedings. All references should be to the most pertinent, up-to-date sources **and the minimum of references** are **25 entries** (for original research paper) and **50 entries** (for review/survey paper). References are written in **IEEE style**. For more complete guide can be accessed at (http://ipmuonline.com/guide/refstyle.pdf). Use of a tool such as **EndNote**, **Mendeley**, or **Zotero** for reference management and formatting, and choose **IEEE style**. Please use a consistent format for references-see examples (8 pt):

[1] Journal/Periodicals

Basic Format:

J. K. Author, "Title of paper," *Abbrev. Title of Journal/Periodical*, vol. *x*, no. *x*, pp. *xxx-xxx*, Abbrev. Month, year, doi: *xxx*.

Examples:

- M. M. Chiampi and L. L. Zilberti, "Induction of electric field in human bodies moving near MRI: An efficient BEM computational procedure," *IEEE Trans. Biomed. Eng.*, vol. 58, pp. 2787–2793, Oct. 2011, doi: 10.1109/TBME.2011.2158315.
- R. Fardel, M. Nagel, F. Nuesch, T. Lippert, and A. Wokaun, "Fabrication of organic light emitting diode pixels by laser-assisted forward transfer," *Appl. Phys. Lett.*, vol. 91, no. 6, Aug. 2007, Art. no. 061103, doi: 10.1063/1.2759475.

[2] Conference Proceedings

Basic Format:

J. K. Author, "Title of paper," in *Abbreviated Name of Conf.*, (location of conference is optional), year, pp. *xxx–xxx*, doi: *xxx*.

Examples:

- G. Veruggio, "The EURON roboethics roadmap," in *Proc. Humanoids '06: 6th IEEE-RAS Int. Conf. Humanoid Robots*, 2006, pp. 612–617, doi: 10.1109/ICHR.2006.321337.
- J. Zhao, G. Sun, G. H. Loh, and Y. Xie, "Energy-efficient GPU design with reconfigurable in-package graphics memory," in *Proc. ACM/IEEE Int. Symp. Low Power Electron. Design (ISLPED)*, Jul. 2012, pp. 403–408, doi: 10.1145/2333660.2333752.

[3] Book

Basic Format:

J. K. Author, "Title of chapter in the book," in *Title of His Published Book*, X. Editor, Ed., *x*th ed. City of Publisher, State (only U.S.), Country: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx*.

Examples:

- A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* in Computational Electrodynamics II, vol. 3, 2nd ed. Norwood, MA, USA: Artech House, 1996.
- R. L. Myer, "Parametric oscillators and nonlinear materials," in *Nonlinear Optics*, vol. 4, P. G. Harper and B. S. Wherret, Eds., San Francisco, CA, USA: Academic, 1977, pp. 47–160.

[4] M. Theses (B.S., M.S.) and Dissertations (Ph.D.)

Basic Format:

J. K. Author, "Title of thesis," M.S. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

J. K. Author, "Title of dissertation," Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

Examples:

- J. O. Williams, "Narrow-band analyzer," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, USA, 1993.
- N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

*In the reference list, however, list all the authors for up to six authors. Use *et al.* only if: 1) The names are not given and 2) List of authors more than 6. *Example*: J. D. Bellamy *et al*., Computer Telephony Integration, New York: Wiley, 2010.

See the examples:

REFERENCES

[1] T. S. Ustun, C. Ozansoy, and A. Zayegh, "Recent developments in microgrids and example cases around the world—A review," *Renew. Sustain. Energy Rev.*, vol. 15, no. 8, pp. 4030–4041, Oct. 2011, doi: 10.1016/j.rser.2011.07.033.

[2] D. Salomonsson, L. Soder, and A. Sannino, "Protection of Low-Voltage DC Microgrids," *IEEE Trans. Power Deliv.*, vol. 24, no. 3, pp. 1045–1053, Jul. 2009, doi: 10.1109/TPWRD.2009.2016622.

[3] S. Chakraborty and M. G. Simoes, "Experimental Evaluation of Active Filtering in a Single-Phase High-Frequency AC Microgrid," *IEEE Trans. Energy Convers.*, vol. 24, no. 3, pp. 673–682, Sep. 2009, doi: 10.1109/TEC.2009.2015998.

[4] S. A. Hosseini, H. A. Abyaneh, S. H. H. Sadeghi, F. Razavi, and A. Nasiri, "An overview of microgrid protection methods and the factors involved," *Renew. Sustain. Energy Rev.*, vol. 64, pp. 174–186, Oct. 2016, doi: 10.1016/j.rser.2016.05.089.

[5] S. Chen, N. Tai, C. Fan, J. Liu, and S. Hong, "Sequence-component-based current differential protection for transmission lines connected with IIGs," *IET Gener. Transm. Distrib.*, vol. 12, no. 12, pp. 3086–3096, Jul. 2018, doi: 10.1049/iet-gtd.2017.1507.

[6] S. Parhizi, H. Lotfi, A. Khodaei, and S. Bahramirad, "State of the Art in Research on Microgrids: A Review," *IEEE Access*, vol. 3, pp. 890–925, 2015, doi: 10.1109/ACCESS.2015.2443119.

[7] S. Chowdhury, S. P. Chowdhury, and P. Crossley, *Microgrids and Active Distribution Networks*. Institution of Engineering and Technology, 2009.

[8] R. Ndou, J. I. Fadiran, S. Chowdhury, and S. P. Chowdhury, "Performance comparison of voltage and frequency based loss of grid protection schemes for microgrids," in *2013 IEEE Power & Energy Society General Meeting*, 2013, pp. 1–5, doi: 10.1109/PESMG.2013.6672788.

[9] S. Liu, T. Bi, A. Xue, and Q. Yang, "Fault analysis of different kinds of distributed generators," in *2011 IEEE Power and Energy Society General Meeting*, Jul. 2011, pp. 1–6, doi: 10.1109/PES.2011.6039596.

[10] K. Jennett, F. Coffele, and C. Booth, "Comprehensive and quantitative analysis of protection problems associated with increasing penetration of inverter-interfaced DG," in *11th IET International Conference on Developments in Power Systems Protection (DPSP 2012)*, 2012, pp. P31–P31, doi: 10.1049/cp.2012.0091.

[11] P. T. Manditereza and R. Bansal, "Renewable distributed generation: The hidden challenges – A review from the protection perspective," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 1457–1465, May 2016, doi: 10.1016/j.rser.2015.12.276.

[12] D. M. Bui, S.-L. Chen, K.-Y. Lien, Y.-R. Chang, Y.-D. Lee, and J.-L. Jiang, "Investigation on transient behaviours of a uni-grounded low-voltage AC microgrid and evaluation on its available fault protection methods: Review and proposals," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 1417–1452, Aug. 2017, doi: 10.1016/j.rser.2016.11.134.

[13] T. N. Boutsika and S. A. Papathanassiou, "Short-circuit calculations in networks with distributed generation," *Electr. Power Syst. Res.*, vol. 78, no. 7, pp. 1181–1191, Jul. 2008, doi: 10.1016/j.epsr.2007.10.003.

[14] H. Margossian, G. Deconinck, and J. Sachau, "Distribution network protection considering grid code requirements for distributed generation," *IET Gener. Transm. Distrib.*, vol. 9, no. 12, pp. 1377–1381, Sep. 2015, doi: 10.1049/iet-gtd.2014.0987.

[15] O. Núñez-Mata, R. Palma-Behnke, F. Valencia, A. Urrutia-Molina, P. Mendoza-Araya, and G. Jiménez-Estévez, "Coupling an adaptive protection system with an energy management system for microgrids," *Electr. J.*, vol. 32, no. 10, p. 106675, Dec. 2019, doi: 10.1016/j.tej.2019.106675.

[16] M. Brucoli and T. C. Green, "Fault behaviour in islanded microgrids," in *Proceedings of the 19th international conference on electricity distribution, CIRED*, 2007, pp. 0548-(1-4).

[17] I. K. Tarsi, A. Sheikholeslami, T. Barforoushi, and S. M. B. Sadati, "Investigating impacts of distributed generation on distribution networks reliability: A mathematical model," in *Proceedings of the 2010 Electric Power Quality and Supply Reliability Conference*, Jun. 2010, pp. 117–124, doi: 10.1109/PQ.2010.5550010.

[18] L. K. Kumpulainen and K. T. Kauhaniemi, "Analysis of the impact of distributed generation on automatic reclosing," in *IEEE PES Power Systems Conference and Exposition, 2004.*, pp. 1152–1157, doi: 10.1109/PSCE.2004.1397623.

[19] A. A. Memon and K. Kauhaniemi, "A critical review of AC Microgrid protection issues and available solutions," *Electr. Power Syst. Res.*, vol. 129, pp. 23–31, Dec. 2015, doi: 10.1016/j.epsr.2015.07.006.

[20] H. A. Abdel-Ghany, A. M. Azmy, N. I. Elkalashy, and E. M. Rashad, "Optimizing DG penetration in distribution networks concerning protection schemes and technical impact," *Electr. Power Syst. Res.*, vol. 128, pp. 113–122, Nov. 2015, doi: 10.1016/j.epsr.2015.07.005.

[21] S. Chaitusaney and A. Yokoyama, "An Appropriate Distributed Generation Sizing Considering Recloser-Fuse Coordination," in *2005 IEEE/PES Transmission & Distribution Conference & Exposition: Asia and Pacific*, pp. 1–6, doi: 10.1109/TDC.2005.1546838.

[22] H. H. Zeineldin, Y. A.-R. I. Mohamed, V. Khadkikar, and V. R. Pandi, "A Protection Coordination Index for Evaluating Distributed Generation Impacts on Protection for Meshed Distribution Systems," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1523–1532, Sep. 2013, doi: 10.1109/TSG.2013.2263745.

[23] D. Eltigani and S. Masri, "Challenges of integrating renewable energy sources to smart grids: A review," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 770–780, Dec. 2015, doi: 10.1016/j.rser.2015.07.140.

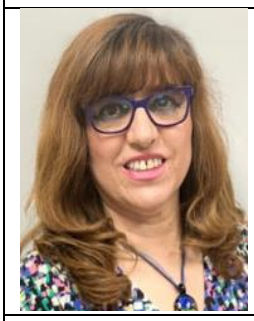
[24] M. M. Eissa (SIEEE), "Protection techniques with renewable resources and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1645–1667, Dec. 2015, doi: 10.1016/j.rser.2015.08.031.

[25] A. Oudalov *et al.*, "Novel Protection Systems for Microgrids," 2009. [Online]. Available: http://www.microgrids.eu/documents/688.pdf.

BIOGRAPHIES OF AUTHORS (10 PT)

The recommended number of authors is at least 2. One of them as a corresponding author.

Please attach clear photo (3x4 cm) and vita. Example of biographies of authors:

	**Abdel-Rahman Hedar** holds a Doctor of Informatics degree from Kyoto University, Japan in 2004. He also received his B.Sc. and M.Sc. (Mathematics) from Assiut University, Egypt in 1993 and 1997, respectively. He is currently an associate professor at Computer Science Department in Jamoum, Umm Al-Qura University, Makkah, Saudi Arabia. He is also an associate professor of artifcial intelligence in Assiut University since January 2012. His research includes meta-heuristics, global optimization, machine learning, data mining, bioinformatics, graph theory and parallel programming. He has published over 70 papers in international journals and conferences. From July 2005 to July 2007, he was a JSPS research fellow in Kyoto University, Japan. He can be contacted at email: ahahmed@uqu.edu.sa or hedar@aun.edu.eg.
	**Patricia Melin** received the D.Sc. degree (Doctor Habilitatus D.Sc.) in computer science from the Polish Academy of Sciences, Warsaw, Poland, with the Dissertation "Hybrid Intelligent Systems for Pattern Recognition using Soft Computing". She is a Professor of Computer Science in the Graduate Division, Tijuana Institute of Technology, Tijuana, Mexico since 1998. In addition, she is serving as Director of Graduate Studies in computer science and Head of the research group on Computational Intelligence (2000–present). Her research interests are in Type-2 Fuzzy Logic, Modular Neural Networks, Pattern Recognition, Neuro-Fuzzy and Genetic-Fuzzy hybrid approaches., She is currently the President of Hispanic American Fuzzy Systems Association (HAFSA) and is the founding Chair of the Mexican Chapter of the IEEE Computational Intelligence Society. She can be contacted at email: pmelin@tectijuana.mx.
	**Dr. Kennedy Okokpujie** holds a Bachelor of Engineering (B.Eng.) in Electrical and Electronics Engineering, Master of Science (M.Sc.) in Electrical and Electronics Engineering, Master of Engineering (M.Eng.) in Electronics and Telecommunication Engineering and Master of Business Administration (MBA), Ph.D in Information and Communication Engineering, besides several professional certificates and skills. He is currently lecturing with the department of Electrical and Information Engineering at Covenant University, Ota, Ogun State, Nigeria. He is a member of the Nigeria Society of Engineers and the Institute of Electrical and Electronics Engineers (IEEE). His research areas of interest include Biometrics, Artificial Intelligent, and Digital signal Processing. He can be contacted at email: kennedy.okokpujie@covenantuniversity.edu.ng.