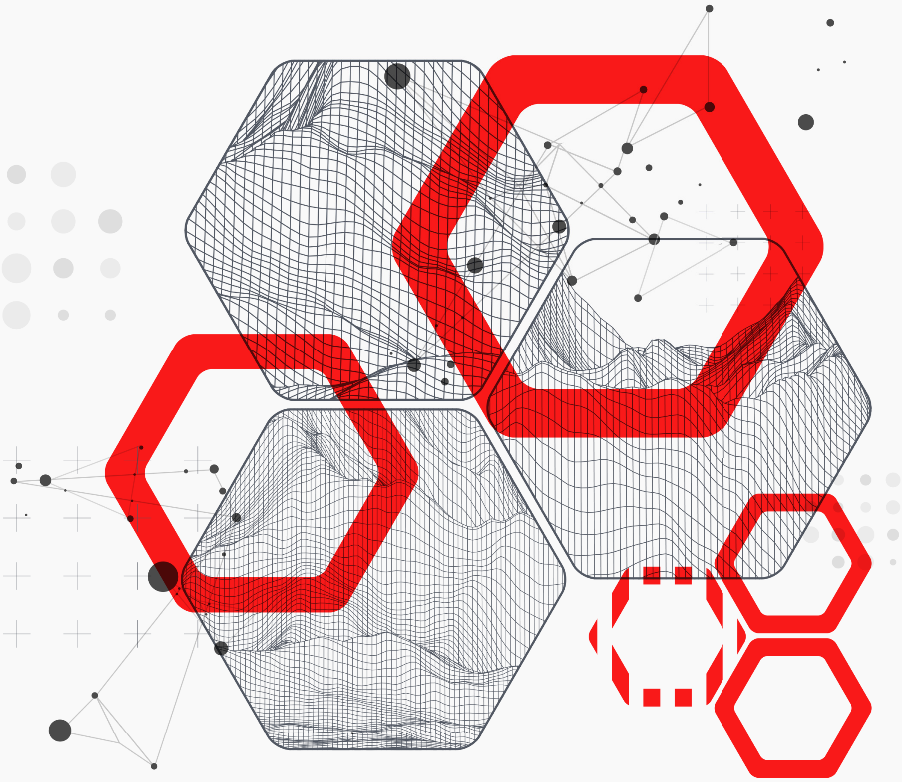
ADCAIJ

Advances in Distributed Computing and Artificial Intelligence Journal



Vol.11 N.3

ADCAIJ.USAL.ES 2022

REGULAR ISSUE

Ediciones Universidad Salamanca

ADCAIJ
Advances in Distributed Computing and Artificial Intelligence Journal

REGULAR ISSUE

Vol. 11 N. 3
2022

Ediciones Universidad Salamanca

eISBN: 2255-2863
Volume 11, number 3
BISITE Research Group
University of Salamanca, 2022

ADVANCES IN DISTRIBUTED COMPUTING AND ARTIFICIAL INTELLIGENCE
https://adcaij.usal.es

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
eISSN: 2255-2863 - DOI: https://doi.org/10.14201/ADCAIJ2022113 - CDU: 004 -
IBIC: Computación e informática (U) - BIC: Computing & Information Technology (U) - BISAC: Computers / General (COM000000)

Regular Issue, Vol. 11, N. 3 (2022)

SCOPE

The Advances in Distributed Computing and Artificial Intelligence Journal (ADCAIJ) is an open access journal that publishes articles which contribute new results associated with distributed computing and artificial intelligence, and their application in different areas.

The artificial intelligence is changing our society. Its application in distributed environments, such as the Internet of Thing (IoT), electronic commerce, mobile communications, wireless devices, distributed computing, Big Data and so on, is increasing and becoming and element of high added value and economic potential in industry and research. These technologies are changing constantly as a result of the large research and technical effort being undertaken in both universities and businesses. The exchange of ideas between scientists and technicians from both academic and business areas is essential to facilitate the development of systems that meet the demands of today's society.

This issue will be focused on the importance of knowledge in advanced digital technologies and their involvement in the different activities of the public-private sector related to specialization in digital technologies and blockchain. The issue also includes articles focusing on research into new technologies and an in-depth look at advanced digital tools from a practical point of view in order to be able to implement them in organisations in peripheral and border areas.

We would like to thank all the contributing authors f or their hard and highly valuable work and members of 0631_DIGITEC_3_E project (Smart growth through the specialization of the cross-border business fabric in advanced digital technologies and blockchain) supported by the European Regional Development Fund (ERDF) through the Interreg Spain-Portugal V-A Program (POCTEP) . Their work has helped to contribute to the success of this issue. Finally, the Editors wish to thank Scientific Committee of Advances in Distributed Computing and Artificial Intelligence Journal for the collaboration of this issue, that notably contributes to improve the quality of the journal.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
eISSN: 2255-2863 - DOI: https://doi.org/10.14201/ADCAIJ2022113 - CDU: 004 -
IBIC: Computación e informática (U) - BIC: Computing & Information Technology (U) - BISAC: Computers / General (COM000000)

Regular Issue, Vol. 11, N. 3 (2022)

INDEX

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 249-261
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.27797

Real-world Human Gender Classification from Oral Region using Convolutional Neural Network

Mohamed Oulad-Kaddour[a], Hamid Haddadou[a], Cristina Conde[b], Daniel Palacios-Alonso[b] and Enrique Cabello[b]

[a] Laboratoire de la Communication dans les Systèmes Informatiques, Ecole naionale Supéerieure en Informatique, BP 68M, 16309, Oued-Smar, Algiers, Algeria
[b] Rey Juan Carlos University, C/Tulipan, s/n,28933, Mostoles, Madrid, Spain
✉ m_ouled_kaddour@esi.dz

KEYWORDS

gender classification; face biometrics; oral region biometrics; convolutional neural networks; deep learning

ABSTRACT

Gender classification is an important biometric task. It has been widely studied in the literature. Face modality is the most studied aspect of human-gender classification. Moreover, the task has also been investigated in terms of different face components such as irises, ears, and the periocular region. In this paper, we aim to investigate gender classification based on the oral region. In the proposed approach, we adopt a convolutional neural network. For experimentation, we extracted the region of interest using the RetinaFace algorithm from the FFHQ faces dataset. We achieved acceptable results, surpassing those that use the mouth as a modality or facial sub-region in geometric approaches. The obtained results also proclaim the importance of the oral region as a facial part lost in the Covid-19 context when people wear facial mask. We suppose that the adaptation of existing facial data analysis solutions from the whole face is indispensable to keep-up their robustness.

1. Introduction

Human-gender classification is a widely studied task. It is one of the most active research areas in biometrics. This is due to the various fields of its usability, such as security, video surveillance, robotic and demographic collection. Human-gender classification describes a binary classification problem. Its study is generally focused on face modality. However, it was also investigated from others modalities, such as fingerprints (Tarare et al., 2015), hands (Affifi, 2019), the periocular region (Busch, 2019), ears (Yaman et al., 2018) and irises (Tapia and Carlos, 2017).

In this paper, an experimental study has been carried out on human-gender classification based on the oral region, specifically on the mouth. Figure 1 shows the parts of the lower face and the targeted region. The scientific literature cited hereafter has been the motivation for our choice of the oral region and of the proposed approach:

- Acceptability and privacy: Naturally, we consider that for any biometric modality targeting a region of interest (ROI) each derived modality obtained by focusing on a sub-region of the original ROI will enhance acceptability. In the case of facial data, even if face modality is sufficiently acceptable by people, we suppose that the oral region is more acceptable in comparison with the whole face. Indeed, people's identities are well hidden and their privacy is protected.
- Eventual alternative for face modality: The oral region can be considered as a biometric modality (Choras´, 2010), there are some scenarios where the lower face or mouth describe the principal part available as in the case of combined occlusion (wearing a hat and black glasses) or an offensive attack. Consequently, the oral region can be an alternative for the whole face modality.
- Modality characteristics: As a biometric modality, the oral region is rich in gender related information and texture. Indeed, the difference in lip angle and format can be observed for the subjects of both gender classes (Figure 2). In addition and for a considerable part of people, it contains gender-reserved synthetic information, such as lipstick for the female class and mustache presence for the male class (Figure 2).

The proposed approach is deployed by adopting deep learning techniques, namely: convolutional neural networks. Indeed, convolutional neural networks or CNNs have been employed in state-of-the-art literature to solve image classification problems because of achieved results (Alzubaidi et al., 2021). In addition, CNNs enable researchers to overcome classical features engineering challenges, as they assure the deployment of powerful solutions in comparison to prior techniques.

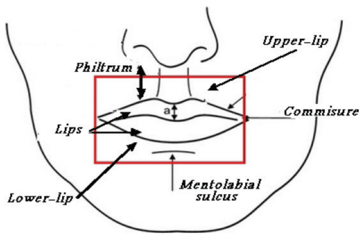


Figure 1. Lower face components

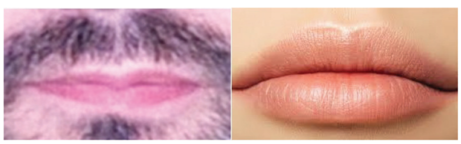


Figure 2. Samples showing some differences between male and female mouth

1.1. Related works

In the literature, the whole human face is the most used modality for facial data analysis and recognition. However, facial components and sub-regions were also used for related tasks. For the gender attribute from the oral region, there are few studies targeting it as a biometric modality (Darryl et al., 2013). In other studies, the oral region was used as a facial sub-region in geometric approaches (Afifi and Abdelhamed, 2019; B. Li, 2012) combining different intermediaries scores.

Darryl Stewart et al. (2013) studied automatic gender classification using static and dynamic features. After face and mouth localization, the authors followed a standard-DCT (discrete cosine transform) based process for feature extraction. Reasonable results were reported for speaker-independent gender classification using the publicly available XM2VTS dataset. Best score values of 82.19% and 18.36% were respectively achieved for both accuracy and EER (Equal Error Rate) metrics.

In Wu et al., 2012, a multi-view gender classification approach was proposed using facial component symmetry. The authors segmented the face according to four facial components, namely: eyes, nose, mouth, and chin. They used the SVM (support vector machine) technique as a sub-classifier for each component. Sum, maximum, product, and fuzzy integral were studied for sub-classifer combinations. An accuracy of 82.23% was obtained for the mouth region.

Bing Li et al. (2012) proposed a framework for human-gender classification by combining facial and external information. They considered five facial regions: forehead, eyes, nose, mouth, and chin with two additional parts: hair and clothing. They exploited LBP (local binary pattern) operator for feature extraction and SVM for classification. The sub-region scores were combined by using various strategies including sum, product, maximum, and majority voting rules. FERET and BCMI face datasets were exploited for experiments. Best classification rates obtained for the region (mouth) were of 82.9% and 83.6% for FERET and BCMI datasets, respectively.

Rai and Khanna (2014) investigated the application of artificial occlusions on the face to perform an occlusion robust system. They defined various occlusion generation strategies by blacking face parts. They used Gabor filters and PCA (principal component analysis). The best classification rate of 85.3% was achieved on FERET face dataset by keeping the lower part of the face via upper face blacking.

Afifi and Abdelhamid (2019) used five isolated facial components including the oral region (mouth) for deep gender classification. The authors used four convolutional neural networks and Adaboost algorithms for final classification for respective sub-facial images: fuzzy face, both eyes, mouth, and nose. The final classification decision was performed by using a linear discriminant classifier. An accuracy of 89.09% was achieved for the mouth region.

Table 1 recapitulates the comparative analysis of the reviewed state-of-the-art literature, by making their principles and best gender classification rate. The oral region was also used to automatize other real-world classification problems. Mouth status estimation (Jie et al., 2016), person identification (Darryl et al., 2013), and lip-reading (Shrestha, 2018) are examples of artificial intelligence tasks already studied from the mouth.

Jie Cao et al. (2016) propose the use of a deep convolutional neural network for mouth status estimation in the wild by taking into consideration various types of attacks. In experimentation, they performed subject-dependent (SI) corresponding to data overlapping inter train and test sets, and subject-independent (SI) to avoid data overlapping inter the train and test sets. Respective accuracies of 90.5% and 84.4% were achieved for both SD and SI experiments.

Table 1. State of the art analysis

Approach	Year	Principle	Best score for the oral region
T.X (Wu et al., 2012)	2012	SVM	82.8%
Darryl (Darryl et al., 2013)	2013	DCT	82.19%
Bing Li (B. Li, 2012)	2012	LBP+SVM	83.6%
Rai (Rai and Khanna, 2014)	2014	Gabor filters + PCA	85.3%
Afifi (Afifi and Abdelhamed, 2019)	2019	CNNs + Adaboost	89.05%

Karan Shrestha (2018) pointed to the complexity of lip reading. The author proposed the training of two deeper separate CNN architectures for real-time word prediction. The Haar classifier was used for face and mouth localization. Batch normalization was applied to speed up the training process and improve stability. A dropout rate of 40% was empirically fixed for the network generalization and overfitting reduction. Best validation accuracy of 77.14% was obtained.

Yannis et al. (2016) proposed LipNet, a recurrent neural network exploiting spatiotemporal convolutions for lip-reading. LipNet is the first model trained with an end-to-end strategy for lip movement translation to text in videos frames. LipNet surpassed prior works and achieved new state-of-the-art accuracy of 95.2% in sentence-level.

In Carrie and Darryl (2020), LipAuth was proposed a system for lip-based authentication for mobile devices. The authors trained convolutional neural networks inspired by LipNet (Yannis et al., 2016). They used the XM2VTS dataset and collected new datasets, qFace and FAVLIPS. An equal error rate of 1.65% was obtained by benchemarrking the system on the XM2VTS.

2. Methodology

In this section, the proposed methodology is presented along with the applied convolutional neural network.

2.1. Overview

The overview of our approach is illustrated in Figure 3. At first we extracted the facial annotations corresponding to left and right eyes, nose, left and right commissure of the mouth. We computed the five facial annotations with the RetinaFace algorithm (Jiankang et al., 2019). Secondly, we performed the localization of the oral region by exploiting the left and right commissure. Finally, we passed the extracted region of interest to the convolutional neural network (CNN) for processing and decision making.

2.2. Convolutional neural network

Convolutional neural networks or CNNs are a powerful learning based technique for image classification. A CNN principally integrates the following layer types: convolutional, pooling, batch-normalization, dropout, fully connected.

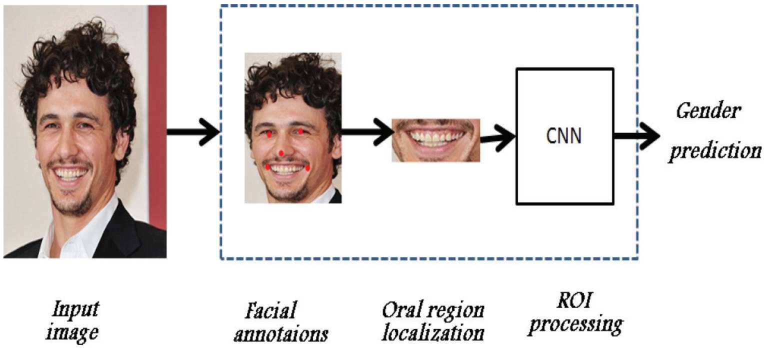


Figure 3. Overview of the proposed methodology

- Convolutional layer: It is a fundamental layer type in a CNN. At this stage, convolutional operator is applied on the input with the goal of learning features to enable detection in future subjects. A convolutional operator is mainly characterized by its kernel size.
- Pooling layer: In this type of layer, pooling operators are applied with the goal of reducing the data- dimensionality. They are generally used following convolutional layers, to generate maps of high resolution. Max-pooling and average-pooling are examples of the most used pooling methods.
- Batch normalization: It is an artificial intelligence technique that aims to speed up the training of very deep neural networks, to maintain stability during training phases.
- Fully connected layers: In a fully connected layer (FC), neurons are connected with all activation's in the previous layers. FC layers are generally placed at the end of CNNs.

Dropout layers: In deep learning, a dropout layer is a layer that aims to avoid the overfitting of trained CNN. A dropout layer simply works by forgetting a part of its input. Dropout layers are generally placed following FCs layers.

In our case, we adopted a sequential convolutional neural network by exploiting various layer types, namely: convolutional (conv2d), batch normalization (BatchNo), maximum pooling (MaxPooling2D), fully connected (Dense), and dropout layers. We used a binary softmax layer for the final decision. The detail of the proposed CNN is shown in Table 2.

3. Experimental setup

The experimental setup is presented in this section by presenting the used dataset, the CNN training conditions and the evaluation metrics.

3.1. Dataset

To carry out the experimentation, we used the FFHQ (Flickr faces high quality) dataset. it is a recent real- world dataset collected for training the StyleGAN (Karras et al., 2019); an adversarial neural network generating realistic artificial fake corpus. FFHQ is a rich dataset in terms of facial variations,

Table 2. Details of the proposed CNN architecture

Layer (type)	**Output Shape**	**Param**
conv2d_1 (Conv2D)	(40, 85, 32)	896
batch_normalization (BatchNo)	(40, 85, 32)	128
max_pooling2d (MaxPooling2D)	(20, 42, 32)	0
conv2d_2 (Conv2D)	(18, 40, 64)	18496
batch_normalization_2	(18, 40, 64)	256
max_pooling2d_2	(9, 20, 64)	0
conv2d_3 (Conv2D)	(7, 18, 128)	73856
max_pooling2d_3	(3, 9, 128)	0
flatten_3 (Flatten)	(3456)	0
batch_normalization	(3456)	13824
FC (Dense)	(256)	884992
batch_normalization	(256)	1024
dropout (Dropout)	(256)	0
batch_normalization	(256)	1024
dense (SoftMax)	(None, 2)	514

Table 3. FFHQ dataset details

	size in faces	size in subjects	Gender-labeling
FFHQ	~70k	~70k	Labeled

such as age, ethnicity, background, facial expressions, and occlusion. It contains 70K faces extracted from images acquired in the wild under high resolution. The dataset was collected from the well-known Flickr website (www.flickr.com/photos/) by selecting facial images under permissive licenses.

Table 3 resumes the FFHQ dataset details by giving its size in faces, its size in subjects and the gender labeling information. Figure 4 shows samples from the FFHQ dataset.

3.2. CNN pre-training and data augmentation

As previously illustrated in Table 2, we fixed the input size at 40*85. For better CNN weights initialization, we firstly assigned small random values. Then, we pre-trained the network by using the dog-cat dataset publicly available on the Kaggle framework. During the network training, we performed a real-time batch augmentation by using the horizontal flip.

According to deep learning techniques' material requirements, all experiments were performed on machine integrating GPU (graphics processing unit) with the goal of speeding up the training time (Table 4). We used 75 epochs in training and we called back the best settings obtained in the intermediary epochs. We used a small batch size of 10 and we experimentally fixed the dropout rate to a portion of 0.25.

Figure 4. FFHQ samples

Table 4. Used GPU memory characteristics

GPU size	GPU type	Specs
12 GO	Nvidia	Telsa K80

3.3. Evaluation metrics

To evaluate the trained network experimentally, we used various metrics based on the following parameters: TP (true positive), TN (true negative), FP (false positive) and FN (false negative). The following metrics were used as performance indicators: accuracy (ACC), the rate of true positive (RTP) in an objective class, equal error rate (EER) and receiver operating characteristic (ROC). The accuracy describes the rate of correctly classified subjects in the whole test-set. It is defined as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

The rate of true positive describes the rate of correctly classified intra-class subjects. It is computed as following for a given targeted class:

$$RTP = \frac{TP}{TP + FP} \tag{2}$$

The equal error rate is determined by computing medium value for optimal values FAR_opt and FRR_opt where FAR=FP/(FP+TN) and FRR=FN/(TP+FN). The EER formula is:

$$EER = \frac{FAR_{opt} + FRR_{opt}}{2} \tag{3}$$

The receiver operating characteristic is a curve that allows a graphical interpretation for the CNN comportment in front of test-set. It is addressed by displaying the evolution of FAR and FRR values during the test phase.

4. Results and discussion

In this section, the obtained experimental results for the proposed approach, using the cited metrics, are presented and discussed. Feature visualisation is performed and a baseline comparison with prior works is established.

4.1. ACC, RTP

For the training of the network, we used a gender-balanced set of 5k images from the FFHQ dataset. For the test, we randomly selected a gender-balanced set of 1000 images. All the images correspond to the extracted oral region from frontal and semi-profile faces. The facial variations of the used sets are recapitulated in Table 6 in terms of facial expression, face pose, image quality, age, and ethnicity. Training and test set details are resumed in Table 5.

The obtained confusion matrix is shown in Table 7. Table 8 resumes the achieved accuracy (ACC) and RTPs (FRP and MRTP) describe the rate of correctly classified subject intra-class. An accuracy of 92.70% has been achieved for the whole test-set. For female and male genders, we got respective values of 94.00% ad 91.40% for the FTP and MTP parameters. It has been noted that the male gender easier to detect than the female gender. It can be justified by the fact that the upper-lip texture for the male gender, is easily detected by the trained convolutional neural network, especially in the case of subjects who have a mustache.

4.2. ROC-AUC, EER

Figure 5 shows the ROC curve traced for the tested network. Table 9 shows the EER error and the AUC (area under the curve) describing the area covered by the ROC curve. By looking at the ROC curve we note clearly the convergence of the classifier as a good discriminator. Moreover, a quantitative value of 0.966 closer to 1 was obtained for the AUC parameter and a high area was covered. In

Table 5. Train and test size

	Female	**Male**	**totale**
Train (in kilo images)	5k	5k	10k
Test	500	500	1000

Table 6. Train and test set's facial variations

Variation	Observation
Facial expression	Random
Face pose	Frontal and semi-profile
Image quality	Acceptable
Age	Random, except young
Ethnicity	Random: Black, White, Asiatic,...

Table 7. Confusion matrix

		Predicted class	
		Female	**Male**
Real class	Female	457	43
	Male	30	470

Table 8. Obtained ACC, FTP and MTP

	ACC	FTP	MTP
FFHQ	92.70%	91.40%	94.00%

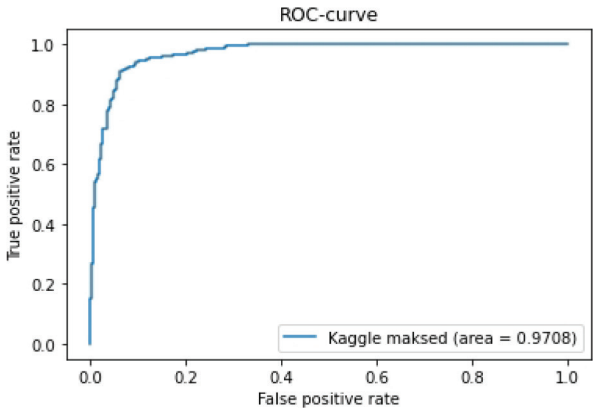


Figure 5. ROC curve

Table 9. Obtained AUC, EER (in probabilistic info.)

Parameter	Value
EER	0.0798
AUC	0.9708

Addition, a low EER error of 0.103 has been achieved. This allows to assert that the human gender can normally be predicted from the oral region.

Figure 6 shows samples of correctly classified and misclassified subjects from the test set. Some factors affecting the classification results have been noted by looking at the whole set of misclassified subjects: facial expressions and upper lip texture. Indeed, we observed that the classifier is more sensitive for smiling faces as a lot of misclassified subjects are smiling. For the second factor of upper lip texture, it has been noted that almost all of the misclassified subjects from the male class are mustacheless, where the visual textures between both gender classes are close. For the last factor, which was

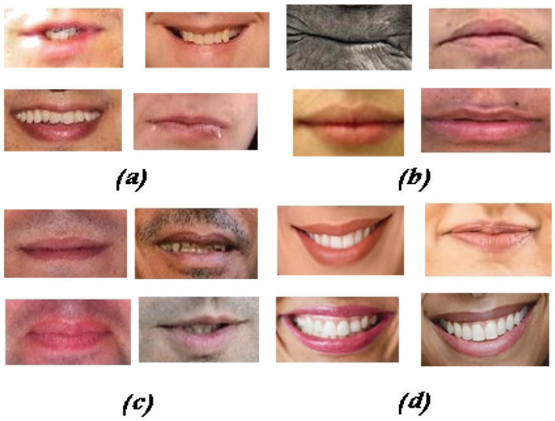


Figure 6. Example of correctly classified and misclassified subjects (a): misclassified males (b): misclassified females (c): correctly classified males (d): correctly classified females

age, it was found that elderly persons were misclassified and this can be justified by the fact that lip format and texture are lost with age. In addition, we observed that the face pose and ethnic variations do have any significant influence over the prediction rates.

4.3. Features visualisation

With the goal of detecting the most discriminative parts of the trained convolutional neural network in the task of gender classification, we used the Grad-Cam technique (Selvaraju et al., 2017). We computed activation maps by applying the Grad-Cam (Selvaraju et al., 2017) technique on the last convolutional operator of the network.

We computed the Grad-Cam class activation maps on several subjects from both female and male classes. Figure 7 shows samples of the obtained results. By looking at the obtained maps, it was found that lip texture and commissure were the parts of the oral region that strongly determined the networks' classification in the case of females. In turn, it was noted that upper and lower lip textures strongly determined the networks' classification in the case of males.

4.4. Baseline comparison

Finally, after evaluating the proposed approach, we made a baseline comparison with prior works using the oral region related part as a biometric modality or facial region. The comparative study presented in Table 10, shows the best-achieved accuracy in each research. It must be noted that in almost all of the existing works the authors experimented on a limited size test set.

As can be seen in Table 10, the gender classification rate obtained in our study surpassed those achieved in the literature. Indeed, we achieved a greater classification rate for the human gender attribute prediction from the oral region using a large, balanced test set of 1000 images derived from a real-world dataset.

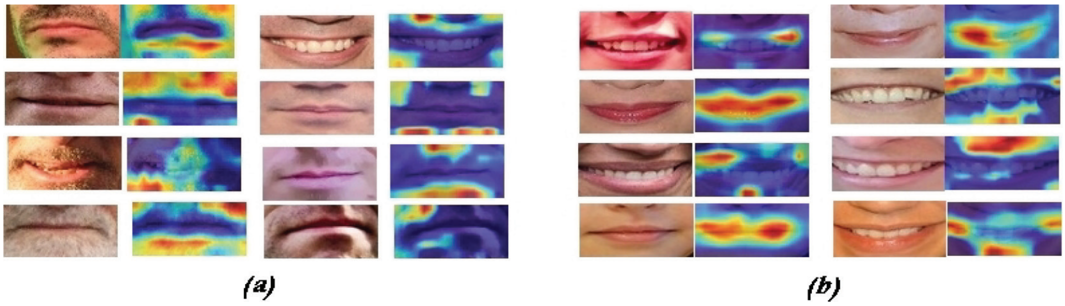


Figure 7. Example of features visualisation for both gender classes (a): male samples (b):female samples

Table 10. Baseline comparison

Approach	**Observation**	**Best score on oral region**
Bing Li (B. Li, 2012)	Use mouth as facial part in geometric approach	83.6%
T. X (Wu et al., 2012)	Use mouth as facial part in geometric approach	82.8%
Darrryl (Darryl et al., 2013)	Gender classification from mouth	82.19%
Rai (Rai and Khanna, 2014)	Use lower face based on mouth	85.3%
Afifi (Afifi and Abdelhamed, 2019)	Use mouth as facial part in geometric approach	89.05%
Proposed	Gender classification from mouth	92.70%

5. Conclusion

In this paper, we proposed a deep learning-based approach for human gender classification from the oral region. For the extraction of the region of interest, the RetinaFace algorithm has been used. For feature extraction and classification, we adopted a convolutional neural network (CNN). CNNs are a powerful tools for image processing and the present implementation is in line with this field of research. To perform our experimentation we used facial images of frontal regular and semi-profile from the real-world FFHQ dataset. The use of semi- profile images widens the field of application. In the present system not only frontal images can be used, but also face variations are incorporated from the beginning, allowing the user to pose in a more natural way.

To assess the effectiveness of our proposed approach, we exploited relevant metrics. Good results were achieved as a low EER was obtained and a global accuracy of 92.70% was returned for a test of 1000 images. Experimental results proclaim the feasibility of using the oral region as a biometric modality and show its importance as a facial part for human gender prediction. We also observed that from the oral region, male gender is easier to detect than the female gender. In relation to the oral features that strongly determined the CNN's classification of gender, lip textures and commissure were important in the female class while upper lip and lower lip textures were more important in the male class. Finally, we evidenced the robustness of the proposed approach by performing a comparative study of prior existing solutions for gender classification from the oral region.

Finally, considering the results of this study, future lines of research should focus on:

- The preparation of a facial dataset related to upper-occlusions with the goal of comparing, experimentally, the results of the proposed approach with the scenario of using the whole human face under upper occlusions.
- The adaptation of the approach for automatic upper-face occlusion detection by integrating a pre-processing step based on deep learning techniques with the goal of enhancing results.
- The generalisation of the approach for the investigation of others categorisation tasks from the oral region, such as: age estimation, race classification and prediction of facial expression.

References

Affifi, M., 2019. 11K hands: Gender recognition and biometric identification using a large dataset of hand images. In *Multimedia Tools and Applications*.

Afifi, M., and Abdelhamed, A., 2019. AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces. In *Journal of Visual Communication and Image Representation*. Elsivier.

Alzubaidi, L., Zhang, J., Humaidi, A., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M., Al-Amidie, M., and Laith, F., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In *J Big Data*. Springer.

Carrie, W., and Darryl, W. S., 2020. Understanding visual lip-based biometric authentication for mobile devices. In *EURASIP J. on Info. Security*.

Choraś , M., 2010. The lip as a biometric. Pattern Anal Applic 13, 105–112. https://doi.org/10.1007/s10044-008-0144-8.

Darryl, S., Adrian, P., and Jianguo, Z., 2013. Gender classification via lips: static and dynamic features. In *IET Biometrics*.

Jiankang, D., Jia, G., Yuxiang, Z., Jinke, Y., Irene, K., and Stefanos, Z., 2019. RetinaFace: Single-stage Dense Face Localisation in the Wild. https://doi.org/10.48550/arXiv.1905.06641.

Jie, C., Haiqing, L., Zhenan, S., and Ran, H., 2016. Accurate mouth state estimation via convolutional neural networks. In *IEEE International Conference on Digital Signal Processing (DSP)*. IEEE.

Karras, T., Laine, S., and Aila, T., 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410. IEEE/CVF.

Li B., Lian X. C., and Lu B. L., 2012. Gender classification by combining clothing, hair and facial component classifiers. In *Neurocomputing*. Elsivier.

Rai, P., and Khanna, P., 2014. A gender classification system robust to occlusion using Gabor features based (2D)2PCA. In *J. Vis. Commun. Image R*.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626. IEEE.

Shrestha, K., 2018. Lip Reading using Neural Network and Deep learning. https://doi.org/10.48550/arXiv.1802.05521.

Tapia, J., and Carlos, A., 2017. Gender Classification from NIR Iris Images Using Deep Learning. In *Deep Learning for Biometrics*. Springer.

Tarare, S., Anjikar, A., and Turkar, H., 2015. Fingerprint Based Gender Classification Using DWT Transform. In *International Conference on Computing Communication Control and Automation*.

Viedma, I., Tapia, J., Iturriaga, A., and Busch, C., 2019. Relevant features for gender classification in NIR periocular images. In *IET Biom.*, page 340–350.

Wu, T. X., Lian, X. C., and Lu, B. L., 2012. Multi-view gender classification using symmetry of facial images. In *Neural Comput. Appl.*

Yaman, D., Eyiokur, F. I., Sezgin, N., and Ekenel, H. K., 2018. Age and gender classification from ear images. In *In International Workshop on Biometrics and Forensics*. IEEE.

Yannis, M. A., Brendan, S., Shimon, W., and Nando, d. F., 2016. LipNet: End-to-End Sentence-level Lipreading. https://doi.org/10.48550/arXiv.1611.01599.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 263-283
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.27822

Learning Curve Analysis on Adam, Sgd, and Adagrad Optimizers on a Convolutional Neural Network Model for Cancer Cells Recognition

José David Zambrano Jara[a] and Sun Bowen[b]

[a, b] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China
✉ macwolfz@gmail.com, 1456174486@qq.com

KEYWORDS

health; medicine; cancer; detection; deep learning; technology; artificial intelligence; machine learning; image classification

ABSTRACT

Is early cancer detection using deep learning models reliable? The creation of expert systems based on Deep Learning can become an asset for the achievement of an early detection, offering a preliminary diagnosis or a second opinion, as if it were a second specialist, thus helping to reduce the mortality rate of cancer patients. In this work, we study the differences and impact of various optimizers and hyperparameters in a Convolutional Neural Network model, to then be tested on different datasets. The results of the tests are analyzed and an implementation of a cancer classification model is proposed focusing on the different approaches of the selected Optimizers as the best method for the achievement of optimal results in accurately improving the detection of cancerous cells. Cancer, despite being considered one of the biggest health problems worldwide, continues to be a major problem because its cause remains unknown. Regular medical check-ups are not frequent in countries where access to specialized health services is not affordable or easily accessible, leading to detection in more advanced stages when the symptoms are quite visible. To reduce cases and mortality rates ensuring early detection is paramount.

1. Introduction

Convolutional Neural Networks (CNNs) have proven useful and have shown highly promising results in Computer-Aided Detection (CADe). The use of Deep Neural Networks minimizes the need for invasive procedures or testing, as well as prevent errors from missed diagnoses where a trained model acts as a second specialist observer, double-checking the results with the benefit that it can bear massive loads of cases in a matter of seconds.

With each advancement in technology, new methods arise, and old methods are refined, an example being CADe methods (Chan et al., 2020). Old methods are being more fine-tuned reaching new levels of precision and accuracy with the help of Deep Learning (Alzubaidi et al., 2021), while new methods are being developed by people around the world which helps in the development of new algorithms (Ozawa et al., 2020) for different models and architectures.

In recent years, new problems have emerged and new techniques have been developed to counteract (Johnson et al., 2021) said problems. Some problems are often associated with using datasets, such as having corrupted data, and or insufficient files. One way to solve this type of problem is by using Data Augmentation which has been proven to work (Shorten and Khoshgoftaar, 2019) by expanding datasets.

Data Augmentation is only able to help improve the dataset to an extent (Hussain et al., 2017), after that it might require different techniques for it to work (Camdemir et al., 2021) and interpret an unrepresentative dataset (More A., 2016) in case one exists.

In this study, the effects and interactions of parameters and tuning are presented, one of these parameters is the 'Optimizer', which updates the weight (Cheng et al., 2021) during Backpropagation.

The Optimizers chosen for this study were ADAM, ADAGRAD, SGD with momentum, and SGD+ Nesterov. The best result from these four optimizers has been tested further on, tuning its parameters for optimal results, allowing the creation of a more robust Lung Cancer Recognition Model.

1.1. Datasets

The main Dataset that was used in this research was 'Mastorides SM. Lung and Colon Cancer Histopathological Image Dataset', which is a collection of benign and cancerous tissue slides (Borkowski et al., 2019) divided on 5 classes for Lung and Colon cells.

In addition, the collections from two datasets, related to cancerous cells, were tested during this study, namely, the Collection of Textures in Colorectal Cancer Histology 'Kather_texture_2016_image_tiles_5000' (Kather et al., 2016), and Histological Images of Human Colorectal Cancer and Healthy Tissue (Kather et al., 2018) 'NCT-CRC-HE-100K'.

1.2. Problem Approach

To take this approach, the first step was to research about different optimizers as background study, and to select the best options that would be used during this paper, that is, those which are the most suitable, in accordance with the selected database.

After selecting the optimizers, the next step was to prepare the information on a structure valid for the model to use. Initially, the slides from each class were separated into a new subset of Training, Validation, and Test, and the images were pre-processed for their further use to train and feed our model.

Then, while defining the structure of the model, a set of hyperparameters have been tuned to seek the most effective values.

During the research, a set of different optimizers were used as well as various sets of hyperparameters values on the previously mentioned Datasets until optimal results were achieved. Once the model reaches a desirable amount of accuracy and loss, it is able to classify the different types of cells of the dataset, separating them into different folders according to his prediction.

Finally, a confusion matrix was created to display the results of the final model classifying performance after the test, showing a chart according the right predictions and errors on our Dataset.

2. Literature Review

2.1. Connection between Deep Learning and Predictive Medicine

The concept of Deep Learning stands for a new learning paradigm in Machine Learning, where a deep architecture model is created through a set of various consecutive layers, and each layer is specialized on extracting multiple features by applying a series of transformations. The goal of such models is to learn complicated and abstract representations of data such as pixels in an image which enter as an input into the first layer, to consequently send the processed output data as an input (Najafabadi et al., 2015) into its next layer.

In this study we use Convolutional Neural Networks (CNN), a class of artificial neural network, which is known for its high accuracy on Image Classification tasks, and it is proven to be very effective in the field of Preventing Medicine where an early detection of a disease is required for timely treatment, especially for patients diagnosed with cancer.

In some countries where a regular check-up is not necessarily frequent due to its cost, lack of specialized means such as equipment or personnel, they usually fail to detect signs and symptoms of cancer at its early stages. Machine Learning proposes diverse methods for approaching these challenges previously mentioned, revolutionizing Health Care field with the presence of artificial intelligence (AI) and precision medicine.

Patients with rare or unique responses to treatment can be identified easier by mixing high precision methods with the new Artificial Intelligence technologies. AI augments human intelligence by using advanced computation and inference to generate insights, support the system 's reasoning and learning capabilities, and facilitate decision-making process for medical personnel. According to recent literature, translating this convergence into clinical research helps address some of the most difficult challenges facing precision medicine, especially for those in which clinical information from patient symptoms (Johnson et al., 2021), clinical history, and lifestyles facilitate personalized diagnosis and prognosis.

The application of new Machine Learning methods and technologies in the field of medicine allows us to provide and adapt early preventive diagnostics and therapeutic strategies to each patient in an optimal and personalized manner. When implemented in health systems, machine learning algorithms can use these datasets to develop recognition or detection models, which, when implemented in healthcare systems, can aid in improving patient care, reducing diagnostic errors, supporting decision-making, and helping clinicians with Electronic Health Record (EHR) data extraction (Scholte et al., 2016) and documentation.

2.2. Deep Learning for illness Classification, Detection and Segmentation

With the advance of technologies, arise new and more complex algorithms for Classification, Detection and Segmentation methods based on Deep Learning, offering innovative approaches and tools for medicine specialists to assist in detecting cancerous cells, lesions and tumors in patients, allowing for a better diagnosis and treatment.

The effectiveness displayed by Neural Networks for task-specific sophisticated feature learning (Bengio, 2014) has helped to achieve high precision detection of illnesses such as tumor progression, where the affected area can be detected and circled inside a box (detection), or highlighted the contours (Segmentation), and separated according to which type of a disease it belongs to (classification).

In other words, the efficacy of molecular imaging diagnostics for the early diagnosis of cancer can be enhanced and the heavy workload of radiologists and physicians can be reduced, especially when there are subtle pathological changes that cannot be seen visually (Yong Xue et al., 2017).

Other studies have suggested the effectiveness of using Computer-Assisted Diagnostic Systems while undertaking tasks such as Colorectal cancer (CRC) screening (Taghiakbari et al., 2021) and prevention examinations.

During these examinations deep learning algorithms automatically extract diminutive (lesser than 5 mm) and small (lesser than 10 mm) polyp features (Tom et al., 2020). Besides, by implementing visualization Enhance methods such as blue-light imaging, narrow-band imaging (NBI), and i-Scan, it is possible to attain improved and reliable pathology predictions (Taghiakbari et al., 2021) during this process.

The massive advantages that Deep Learning has to offer to the medical field are quite evident, evolving to cover the 3 most important applications for image analysis in the medical field, such as target detection, segmentation, and classification, however, new opportunities also brought along with certain challenges. For example, the use of different Datasets on which deep learning models profoundly rely on, are limited and rather scarce, since the process for training these models require a massive load of data, or even by having large amounts of data, these datasets are still not fit enough for creating optimal and highly accurate models, so different strategies must be applied to improve the Accuracy and Loss of these models, and their inherent tools such as optimizers, which are studied in this research.

3. Optimizers

3.1. Stochastic Gradient Descent (SGD)

The main feature of the optimizer is that unlike the original algorithm, which was based on (Gradient Descent), where a series of calculations are made on the entire dataset, it takes into consideration a small subset of randomly chosen datasets.

As a result, the computing speed increases (Tom et al., 2020) while the storage requirements decrease. SGD updates the weights after seeing each data point instead of the entire dataset, however, it makes rather noisy jumps away from the optimal values since it is influenced by every individual input. The formula for every epoch and every sample on this algorithm is:

$$\theta^1 = \theta^0 - \alpha . \nabla_{\theta} J(\theta) \tag{1}$$

Where θ^1, as an element of the domain of Real numbers represents the next position of the parameter of the objective function $J(\theta)$, which keeps updating through each training example to the opposite direction of the gradient of the objective function $\nabla_\theta J(\theta)$. Being this objective Function the gradient of $J(\theta)$ respect to θ, and α the learning rate, which reverse the direction of this gradient obtaining the direction of maximum descent (Elashmawi, 2019). In other words, As the position of the cost function gets smaller and smaller, the formula recounts the next position towards the steepest descent. (Figure 1)

The training environments of high complexity in which Gradient Descent methods fail to work properly has led to the development of several new algorithms that complement technological advances (Zhou et al., 2020). One disadvantage of SGD is how it evenly scales the gradient in every direction (Keskar and Socher, 2017), turning the process of tuning the learning rates α fairly arduous. Besides, SGD Algorithm convergence rate is rather slow, due to the inherent propagate process that goes backward and forward for every record (Ruder, 2016). Another way of decrease the noise of SGD is to add the concept of Momentum (Bengio, 2012) to the model. The hyperparameters of the model may have the tendency of changing in one direction; with Momentum, the model can learn faster by minimizing the attention paid to details on the few examples that are being shown to it.

$$V = \gamma . \vartheta + \eta . \nabla_\theta J(\theta) \tag{2}$$

$$\theta = \theta - \alpha \vartheta \tag{3}$$

Where a fraction 'γ' (also called 'momentum', generally set to a value around 0.9) of the update vector is added to the current update velocity vector represented by ϑ, in order to attain a faster convergence and reduced oscillation. Calculating $\theta = \theta - \alpha\vartheta$ helps to find the most suitable value by iteratively obtaining the eigenvalues, where α is the drop coefficient, that is, the step size and the learning rate.

In other words, finding the magnitude of each drop. A larger coefficient leads to a greater difference in each calculation, while a smaller coefficient, causes a smaller difference, however, the iterative calculation time is relatively longer. The initial value of θ can be assigned randomly, as for example,

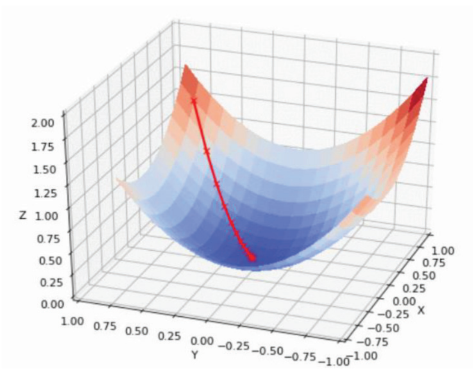


Figure 1. Convex Function of Gradient Descent. (Retrieved February 4, 2022, from https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/)

the initial value can be set to 0. However, choosing to blindly ignore samples because they do not have typical features is reflected in a higher loss; to this end, an acceleration term is added.

$$\vartheta_t = \gamma . \vartheta_{t-1} + \eta . \nabla_\theta J(\theta - \gamma . \vartheta_{t-1}) \tag{4}$$

$$\theta = \theta - \alpha \vartheta \tag{5}$$

Being ϑ_t, the velocity of time t and ϑ_{t-1}, the velocity of the previous time step, and $\eta . \nabla_\theta J(\theta - \gamma . \vartheta_{t-1})$ representing gradient at the particular time 't', and momentum γ is set a value of around 0.9 as well. This new Gradient has the particularity that when is close to the optimal value, or is near to the minimum value of the slope, the result of the gradient becomes negative, directing the gradient update back towards θ*t* (Gylberth, 2018), allowing it to avoid oscillations. Nesterov acceleration presents some advantages over normal SGD with Momentum, where the gradient of future position is taken (Schmidt et al., 2018) instead of current position. In other words, the model gains momentum during training, so when it finds an odd example, because of the added momentum it does not pay too much attention to it.

Discarding it leads to a loss decrease that is not too abrupt as it is with normal SGD + Momentum (Ruder, 2016), this is where the weight updates are decelerated, so they become small again, allowing future examples to fine-tune the current model (Sutskever et al., 2013), updating the weights and bias rather dynamically (Figure 2).

3.2. Adagrad

Adagrad adjusts the learning rate to the parameters, carrying out a set of small updates for recurrently features, and a bigger learning rate for uncommon features. The main idea of this algorithm is to keep in memory the sum of squares of the gradients with respective θ_i parameter up to a certain point.

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{(G_{t,ii} + \varepsilon)}} . G_{t,i} \tag{6}$$

Where $\theta_{t,i}$ is the current parameter value on i during every time step t, while $\theta_{t+1,i}$ is the updated parameter after one time step. η (ETA) value is usually set to 0.01, and $\in$ representing a relatively

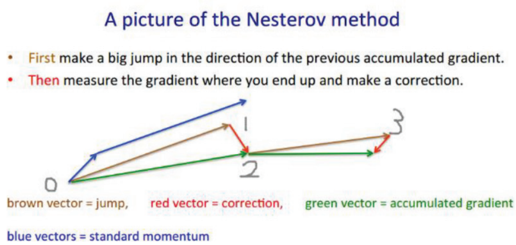


Figure 2. Nesterov method process (Retrieved on February 4, 2022 from https://towardsdatascience.com/gradient-descent-explained-9b953fc0d2c)

small number used to ensure that the denominator is not 0 (usually on the order of 1e-8), G_t contains the gradient estimate at time step 't' with relation to the parameters θ_i (Ruder, 2016), obtained as the partial derivative of the objective function respect to the parameter θ_i, and $G_{t,ii}$ as a diagonal matrix where each diagonal element i,i is the sum of the squares of the gradients respect to θ_i up to time step t.

$$G_{t,ii} = 2{,}66 = \nabla_\theta J(\theta_{t,i}) \tag{7}$$

In other words, the learning rate is divided by the square root of all previously obtained gradients, the momentum concept is removed, while the learning rates are adjusted according to the parameters. This is unlike the other optimizers, where an update is performed for all parameters θ, instead of using a different learning rate for every parameter on each Epoch.

3.3. Adam

This method incorporates the momentum concept from «SGD with momentum» and adaptive learning rate from «Ada delta», as well as the best features of AdaGrad and RMSProp practices, making Adam suitable to handle sparse gradient gradients (Kingma and Ba, 2015) easily on noisy problems. Sparse gradients are, in essence, a way to shift the approach of the Original Past Gradient (OPG) from calculating gradient vectors only at observed points to computing gradients over the entire domain (Ye and Xie, 2012). This algorithm can obtain better results than many other optimizers in the field of deep learning. This was demonstrated (Figure 3) when applied to an analysis using logistic regression (Kingma and Ba, 2015) and CNN on the MNIST dataset and CIFAR-10 dataset respectively.

A crucial aspect of ADAM 's competitive performance is that it can solve practical deep learning problems with large datasets and models while having to go through a minimal tuning (Karpathy, 2017), as well as epochs to achieve such results. Although having such good performance, later works have suggested the possibility of lacking the ability of adaptive methods to outperform SGD (Wilson et al., 2017) when measured by their skill to generalize. This is, that the model can display proper adaptation when new pieces of information are added to it (Brownlee 2021) from the same dataset as the original input.

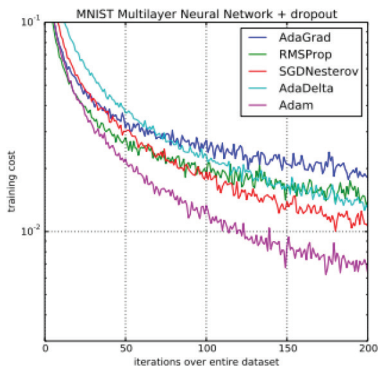


Figure 3. Different optimization methods graph on a Multilayer perceptron training (Retrieved on February 7, 2022, from https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/)

4. Training the network and result analysis

4.1. Lung and Colon Cancer Histopathological Images Dataset

For the present study, the main dataset selected was 'Lung and Colon Cancer Histopathological Images', which was used on a model of CNN based on Xception's algorithm architecture, and tested on different hyperparameters, creating their respective Learning Curves for its analysis. These learning curves represent the performance of the model during its training and validation, in other words, a graph of the model 's learning performance over time.

By watching these graphs, the algorithms can be analyzed during their training and validation process on their dataset respectively, where the plotted lines show how close or far the model's learning, learning or prediction get from the group of samples in the dataset. The goal is to achieve higher accuracy and low loss, when the graph starts to reach a convergence point (Liu et al., 2020) and stabilizes. The optimizers ADAM, ADAGRAD, SGD+momentum, and SGD+Nesterov were respectively tested in the created model with a dataset of 25000 samples, divided into 60% for the Training Process, 20% for the Validation Process, and 20% for the Testing process.

From these Training and Validation tasks, a series of learning curve charts were made, to obtain a better view and easier understanding (Figure 4) from their Accuracy and Loss values (depicted on the y-axis).

For a fair comparison, the models used the same number of iterations (50 per model) called Epochs (depicted on the x-axis), fed by batches of 32 samples each, and a 180x180 pixels rescaling and tested on a model with different unique Optimizers, which each one presented its advantages and shortcomings.

Also, during these testing and comparison early stages, the models were tested with a Dropout of 0.5 which is the recommended value to start (between 0.2 and 0.5). The dropout value selects the percentage of neurons that is to be discarded during the Validation process, so overspecializing on one aspect is avoided and this creates a bias in its predictions, lowering the accuracy of this model.

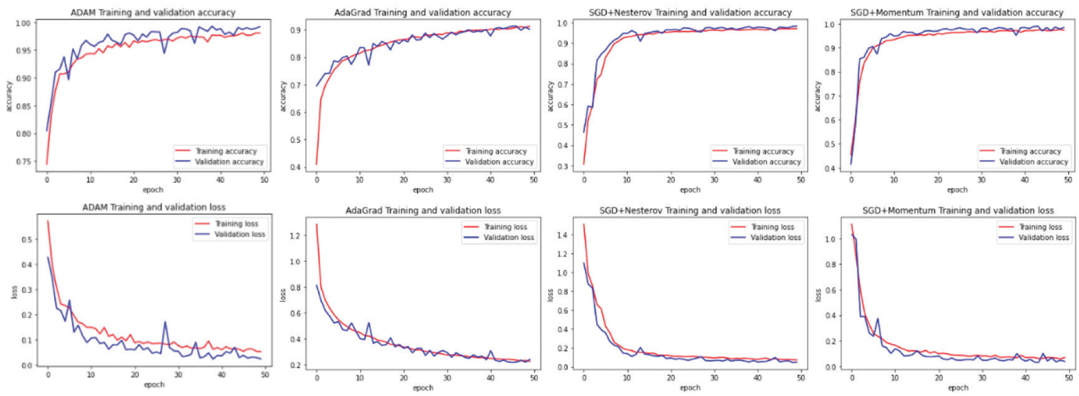


Figure 4. Comparison of learning Curves from Different optimization Algorithms on 'Lung and Colon Cancer Histopathological Images' Dataset (from left to right: ADAM, ADAGRAD, SGD+NESTEROV, SGD)

This dropout technique is not used during the testing, since all the neurons are used here, that is why the accuracy and loss during the training dataset did not show abrupt peaks as the Validation did.

By randomly discarding leaving inactive a percentage of cells during the Validation process, Dropout helps with generalizing the model and avoids the models entering into a state called overfitting. This is made by allowing the model to focus on the different aspects of the features of input samples, creating a type of impartiality on these features to avoid any predisposition that could affect the judgment of the model predictions.

The results of using different optimizers on one dataset for the recognition of Cancerous Lung cells were brought together. It is clear that Adam works best in matters of overall Accuracy and Loss.

Adam optimizer showed satisfactory results, reaching good accuracy in both the Training and Validation process, and the lack of sudden jumps and drifting suggests that the model graph is slowly reaching into convergence, reaching a minimum Validation Accuracy and Loss of 0.9855 and 0.04 respectively, turning this model into a candidate for further tests with a larger number of Epochs and different Dropout values.

Adagrad optimizer performed well on this Dataset, where it reached early acceptable results, a minimum loss value of 0.1286, and a maximum accuracy of 0.9571, both for the Validation process.

However, their results stagnated on later Epochs, preventing the model from further learning on the current Dataset. As stated before, one of the drawbacks with Adagrad is how it monotonically increases the sum of squares of the gradients on further Epochs, leading to decay of the Learning Rate to decay until the parameter does not update anymore, in other words, it stops learning.

As for 'SGD + Momentum', the Learning Curve exhibited relatively slow learning at first, however, it eventually started to improve on a rather stable pattern as it advanced on later Epochs, reaching up to a max accuracy of 0.9744 and min loss value of 0.0876, turning this optimizer version of the model into another candidate for further tests (Zhou et al., 2021), with a larger number of Epochs and a tune of hyperparameters if needed.

Finally, SGD + Nesterov optimization displayed rather stable values, 0.068 for minimum loss and 0.9792 for maximum accuracy, yet not as high as Adam or SGD + Momentum. In this case, the Nesterov optimization method slowed down the learning process preventing the model from both stabilizing and reaching higher accuracy values with a low loss rate.

To corroborate these results, the same optimizers were tested on different datasets, to verify and compare the results with those shown in Figure 4.

4.2. Kather 2016 Dataset

After the training process done on 'Kather_texture_2016_image_tiles_5000' Dataset, the results from the learning curves show the performance of the models (Figure 5), all of them with a Dropout value of 0.5, and separated by the Optimizer that was used on it for the current Dataset in batches of 32 samples, alongside with a rescaling of the size of the image down to 180 x 180 pixels.

On the current dataset results (Figure 5), through the learning curve is observable that ADAM model was presenting the most stable pattern among all the other optimizers.

As expected, ADAM optimizer showed promising results, similar to those on 'Lung and Colon Cancer Histopathological Images' Dataset. In this case, the learning curve for Validation Accuracy showed peaks so high reaching a max 100% accuracy, and an average of 0.968 accuracy during the 50 Epochs, showing a slow tendency to stabilize on every recurrence, without quite finding a convergence point on early Epochs.

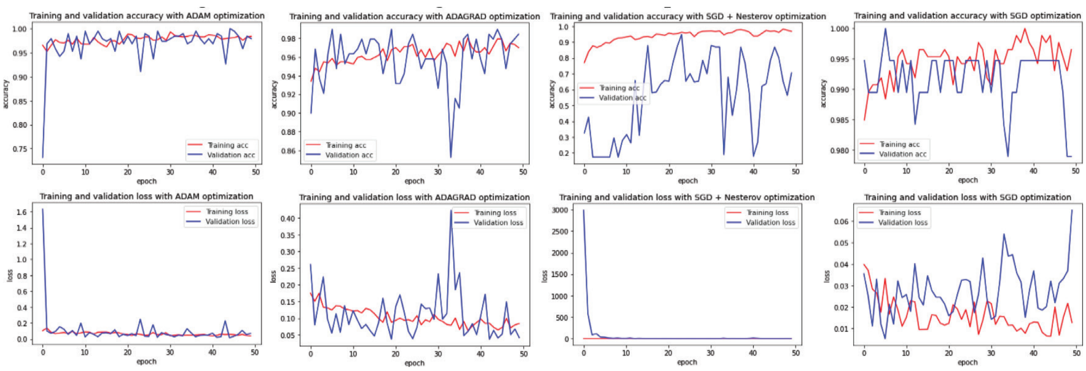


Figure 5. Comparison of learning Curves from Different optimization Algorithms on 'Kather_texture_2016_image_tiles_5000' Dataset (from left to right: ADAM, ADAGRAD, SGD+NESTEROV, SGD)

On the other hand, on the Learning Curve for Loss Function, it is noticeable how the gaps are relatively small. Moreover, the gaps reach a minimum loss of 0.0135, meaning that the model is prone to avoid making mistakes or incorrect predictions during the testing process, which is paramount during cancer detection, affecting greatly the patient's diagnosis.

Also, it is noticeable how Adagrad optimizer, despite having peaks sometimes gets closer to an ideal model, does not get to stabilize. However, a good strategy is seizing this optimizer's skill of reaching higher Accuracy values in early Epochs, in case of encountering learning stagnation problems in later stages. In this case, reaching a minimum loss of 0.0366 and maximum accuracy value of 0.9895 during the Validation process.

On the other hand, 'SGD + Nesterov' did not show as good results as those visible on the previous dataset (Figure 4), however, this optimizer´s Loss Learning Curve remained low and stable. These metrics reveal that this model is not suitable yet to fully extract the features of the samples of this dataset properly, although there is a possibility for improvement of the feature extraction by applying different techniques such as data augmentation in the data, to create bigger a sample set that proposes new opportunities to the model, to obtain more features of each class (Dönicke et al., 2019; Candemir et al., 2021) of the dataset.

Nonetheless, this model reached a maximum accuracy value of 0.9474, which although being acceptable, is not as high as the other versions of this model, and a loss value of 0.1351 which is rather high on a metric whose goal is to be as low as possible.

Lastly, the 'SGD' optimization method displayed a performance full of peaks on this dataset. The momentum concept allows to minimize the attention that is paid from the model to details, this method helps to avoid overfitting while boosting the learning speed.

At first glance, the gaps on the model, especially those close to the 50th Epoch would tell it's straying far away from convergence, but this can also be seen as a result of the momentum, which in this case produced positive Accuracy results, having an average of 0.991774, which is impressively high, a maximum Validation Accuracy of 100, and a loss value of 0.0052

4.3. NCT-CRC-HE-100K Dataset

Finally, the results (Figure 6) confirm how the ADAM optimization method was the one who performed the best at the end of the training phase from the 'NCT-CRC-HE-100K' Dataset, followed by SGD+Nesterov.

Similarly, with the other datasets, the parameters used for these models were batches of 32 samples at the time and a rescaling of 180 x 180 pixels, a Dropout value of 0.5, and a Momentum value of 0,9.

By observing the results obtained from this third dataset, and comparing them with those from the other two datasets, it is noticeable how ADAM maintained its great performance, with a maximum Validation Accuracy of 0.9855 and a minimum Loss of 0.0492, outperforming the other optimizers, such as Adagrad model which presented an Accuracy value average of 0.779, except counted times when hitting a maximum Accuracy value of 0.9518 and Loss rate of 0.1286

'SGD+ Nesterov' showed a good performance, similar to 'SGD with Momentum', both above 97% of Accuracy, and rather low Loss values of 0.0876 and 0.0876 respectively.

After plotting the learning curves from different optimizers, it becomes perceptible how Adam, despite having some peaks, still shows a tendency to a convergence point, displaying a good overall on the different datasets, and for this reason, this optimizer was chosen for further experiments in order to obtain an optimal model with efficient predictions, which is paramount at the moment to detect cancerous cells at early stages saving lives.

The result values during each Training and Validation process are presented in Table 1, divided by each dataset on every tested optimizer, count of how many values during these Epochs did not rise above a threshold of '0.1' for Loss, and those which were able to exceed a value of 0.95.

Table 1 also displays the Maximum Accuracy and Minimum Loss value that each model reached during its Validation, followed by an overall average of Accuracy and Loss for a better understanding of this model behavior towards the selected Dataset, and lastly, the time required (measured in seconds) to complete the whole process.

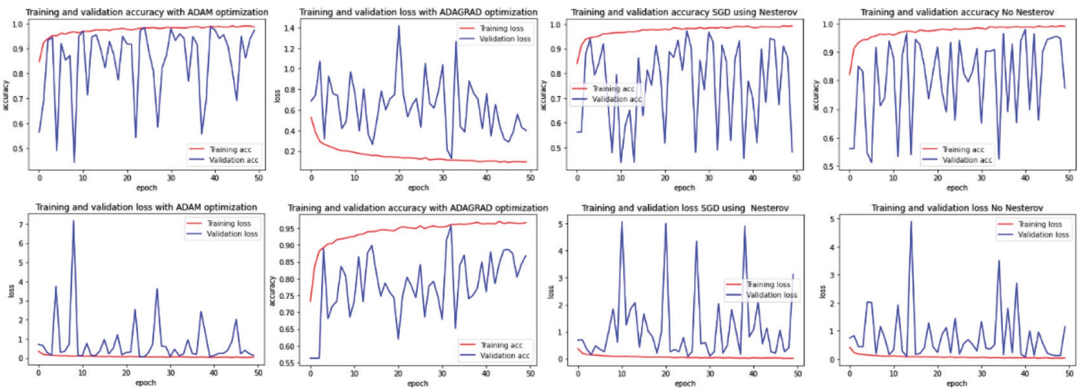


Figure 6. Comparison of learning Curves from Different optimization Algorithms on 'NCT-CRC-HE-100K' Dataset (from left to right: ADAM, ADAGRAD, SGD+NESTEROV, SGD)

Table 1. Performance Results and Comparison for Optimizers on custom model on Kather, CRC-VAL-HE-7K, and LC25000 Datasets

		LOSS	ACC	VAL LOSS	VAL ACC	MIN LOSS	MAX ACC	MIN VAL LOSS	MAX VAL		AVG ACC	AVG VAL LOSS	AVG VAL ACC	Sum of Speed
									ACC	AVG LOSS				
KATHER	ADAM	48	50	39	44	0.0061	0.993	0.0135	1	0.063276	0.97764	0.103282	0.96893	1720
	ADAGRAD	29	47	29	35	0.0624	0.9803	0.0366	0.9895	0.102746	0.96352	0.110252	0.958314	1720
	SGD+Nesterov	12	27	0	0	0.0638	0.9815	0.1351	0.9474	0.174932	0.939612	79.008334	0.561362	1576
	SGD	50	50	50	50	0.0063	1	0.0052	1	0.016194	0.994618	0.027376	0.991774	1733
	ADAM (0.2 Dropout)	79	89	19	34	0.016	0.9954	0.0228	0.9947	0.089106	0.970348	0.679409	0.852684	9297
	ADAM (0.5 Dropout)	1	30	0	4	0.0945	0.9678	0.1226	0.9575	0.172607	0.940304	0.650614	0.8614	4303
CRC-VAL-HE-	ADAM	40	47	4	10	0.0265	0.9913	0.0492	0.9855	0.076972	0.972408	0.773822	0.848666	9696
	ADAGRAD	10	29	0	1	0.0872	0.9718	0.1286	0.9571	0.154862	0.939204	0.629188	0.77994	8158
	ADAM 0.2 DROPOUT	90	96	15	26	0.0061	0.9981	0.0072	0.9965	0.05107	0.981726	0.923743	0.839584	17947
	ADAM (0.5 Dropout)	100	100	31	51	0.0127	0.9969	0.0191	0.9947	0.04181	0.9862	0.42132	0.928086	6501
	SGD+MOMENTUM	41	45	2	2	0.0154	0.9941	0.0876	0.9744	0.071768	0.97306	1.127848	0.752724	7800
LC25000	ADAM	40	47	4	10	0.0265	0.9913	0.0492	0.9855	0.076972	0.972408	0.773822	0.848666	9696
	ADAGRAD	10	29	0	1	0.0872	0.9718	0.1286	0.9571	0.154862	0.939204	0.629188	0.77994	1720
	SGD+Nesterov	41	45	2	5	0.0238	0.9913	0.068	0.9792	0.079746	0.970302	0.80578	0.807776	1576
	SGD	41	45	2	2	0.0154	0.9941	0.0876	0.9744	0.071768	0.97306	1.127848	0.752724	1733
	ADAM (0.2 Dropout)	90	96	15	26	0.0061	0.9981	0.0072	0.9965	0.05107	0.981726	0.923743	0.839584	9297
	ADAM (0.5 Dropout)	88	96	15	27	0.0074	0.9971	0.0434	0.9841	0.052853	0.98102	0.46387	0.889964	4303

These experimental results summarized in Table 1 provide a comparison between optimizers on the proposed classification method on different datasets, where so is observable the improvement between each Optimizer on the current Model, delivering initial results on the projected classification pipeline.

Overall, the model achieved on its performance a peak of 99.8% and 99,7 of accuracy on Adam with 0.2 Dropout and 0.5 Dropout respectively and a loss of 0.006% and 0.0074%. The overall result displayed by the proposed model is able to capture features from cell slides to detect cancerous cells in patients in real-world scene scenarios.

From the table, it was evident which models performed the best, and by diagnosing the properties of their Learning curves, the datasets can be analyzed as well as the model behavior during the fitting process. In this case, sudden jumps and drops shown on the Learning Curves suggest that the validation dataset is unrepresentative

4.4. Unrepresentative Dataset

This definition applies to datasets with a small number of samples compared to a large, representative dataset. Resulting in poor feature extraction by the model that is in training, thus a low accuracy and high error rate on our model, regardless of the quality of it.

In the field of Deep Learning, the use of a high-quality model and a good optimizer alone is not enough to achieve satisfactory results. Nevertheless, the dataset that is employed is a paramount factor for accomplishing the desired outcome.

When choosing a Dataset, there are several factors to have in mind, such as the presence or lack of noise from the samples, and mostly the availability of the data, whether be due to rather scarce data or the lack of public information in existence. The analysis of a problem of this nature has a big impact on choosing an optimal size of samples needed to feed the model. There is no concise number of samples for a perfect sized dataset, since it varies depending on the nature and complexity of the problem, or on the level of robustness of the model at hand. Nevertheless, an answer which fits is likely an impossible task, but it is advised to start from the factor 10 rule for sample size requirements, which means to have a number of samples at least 10 times more (Alwosheel et al., 2018) than the number of parameters.

Even by following the factor 10 rule, the optimal size of samples for the selected dataset might not be achieved. In occasions where the difference between classes is small, a larger selection might be necessary, as for example to work with ImageNet Dataset database (Deng et al., 2009), it is recommended at least 1000 examples per class (Krizhevsky et al., 2012) for training data.

An optimal dataset selection allows the network to achieve a better generalization, which is observable after the model has concluded the training and validation process. The two most common cases that could arise during the fitting process are Unrepresentative Training Dataset, and Unrepresentative Validation Dataset.

4.5. Unrepresentative Training Dataset

The training dataset lacks sufficient information to facilitate learning compared to the validation dataset, which prevents a learner or model from achieving high accuracy and a smaller loss rate.

This situation can be easily identified by a learning curve of training loss that shows improvement and similarly to the learning curve for validation loss that also shows improvement, but a large gap remains between both curves (Figure 7).

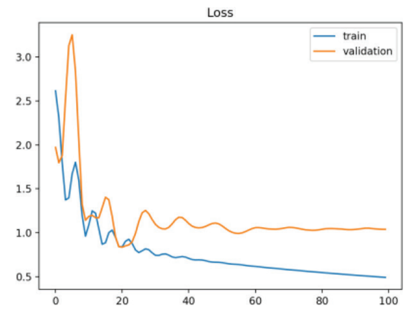


Figure 7. Unrepresentative Training Dataset Learning Curve

Even after applying the factor' 10 rule, when the sample size is too small, a display of rather noisy results on the Learning Curve is obtained, exhibiting a visible inconsistency between train and validation outcome. In such cases, an alternative is to re-do a split of the data samples with the idea that the new examples may be more significant and representative of the model at the moment of extracting its features, as well as applying data augmentation techniques to improve the quality and quantity of the samples provided to the Training Dataset.

4.6. Unrepresentative Validation Dataset

This case occurs when it is not possible to evaluate the generalization ability of the model from the validation dataset, which is to say when there are too many examples from the validation dataset in contrast with the training dataset. This can be observed in a learning curve that appears to fit the training loss well (or other fitting curves) and a learning curve that appears to fit the validation loss well but is characterized by noise. On the learning curves from Loss training, excellent results seem to appear evident, and yet, show unstable changes as is observable on sudden high and low peaks on the validation loss (Figure 8).

In this second example is noticeable the same jumps and drops as those from the learning curve in the model of study, since during the training the data was rather limited and after using different

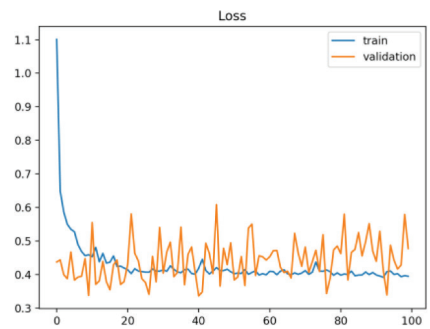


Figure 8. Unrepresentative validation dataset

techniques such as Data Augmentation for increasing the number of samples, the validation data was still unrepresentative compared to the training data.

This model's training was made on a Dropout of 0.5, this means that during the training process (it does not use this during the validation process) 50% of features nodes or neurons are set to 0 because they are not going to be used.

Dropout helps with generalizing the model and avoids the models enter into a state called overfitting.

This dropout technique is not used during the testing, since here all the neurons are used, that is why the accuracy and loss during the dataset training did not show abrupt peaks as the Validation did.

After performing this exploratory testing on Lung Cancer Database, using different Hyperparameters on the best performing optimizer so far (ADAM), the input values for these Dropouts were 0.5 and 0.2 respectively, which after plotting the learning curve, produced similar results, after a larger number of Epochs (Figure 9, Figure 10).

In the current study case (Figure 10) the learning curves showed how a higher Dropout of 0.5 resulted in a rather stable Validation Loss in comparison with 0.2, which makes sense as the model by discarding half of the neurons allow itself to avoid wrong weights or paths easier, but this also has affected the Validation Accuracy where a lower dropout value of 0.2 (Figure 10) exposed better results

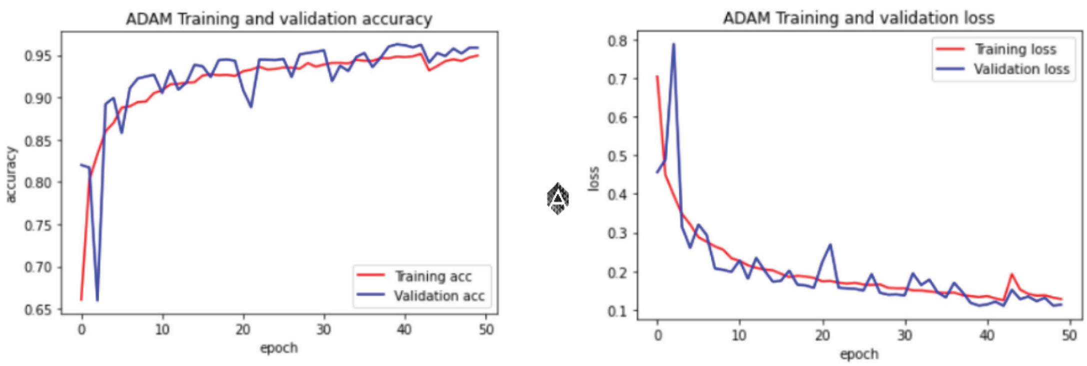


Figure 9. Learning Curve for ADAM optimization with 0.5 Dropout

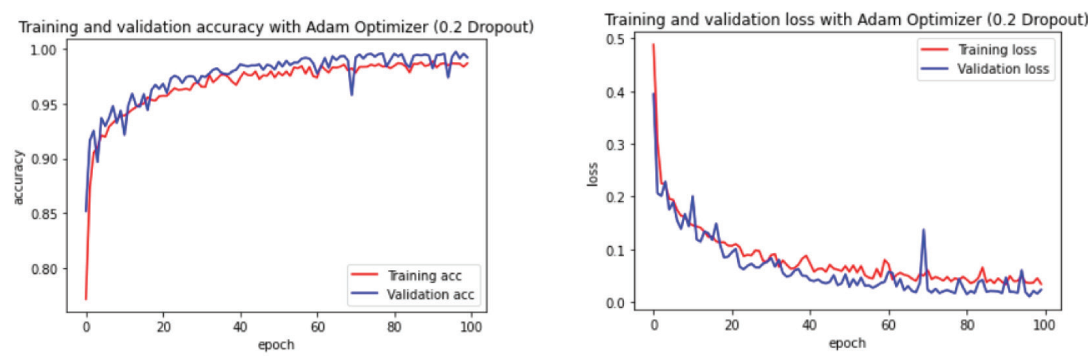


Figure 10. Learning Curve for ADAM optimization with 0.2 Dropout

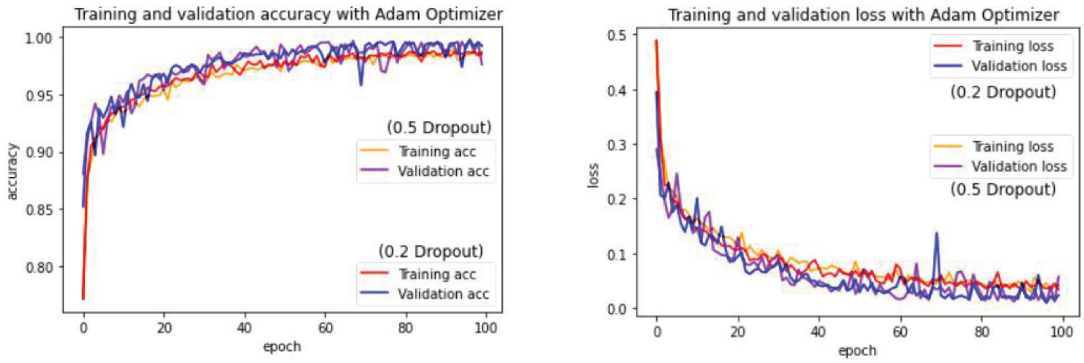


Figure 11. Juxtaposition of learning rates from different Dropout values on ADAM optimization method

on validation accuracy, which is acceptable in this model since it got a positive outcome during the evaluation of this model, delivering a Validation Accuracy of 0.9965 and a Validation loss of 0.0072.

Thus, in this study case, the chosen optimizer for this model was Adam on a Dropout of 0.2 since the overall proved to be better as its observable while transposing the learning curve results as it's shown on Figure 11.

Other studies suggest the possibility a different approach of SGD actually outperforming ADAM in later stages or epochs during the training of the model by Switching from ADAM to SGD (SWATS).

This strategy was tested in cases such as Tiny-ImageNet problem, where the switches from Adam to SGD lead to significant though temporary degradation in their performance. This is, after an abrupt drop from 80% to 52% caused by this switch, the model eventually recovered, achieving an even better peak on testing accuracy compared to Adam, which in some cases lead to a stagnation (Keskar and Socher, 2017) in performance.

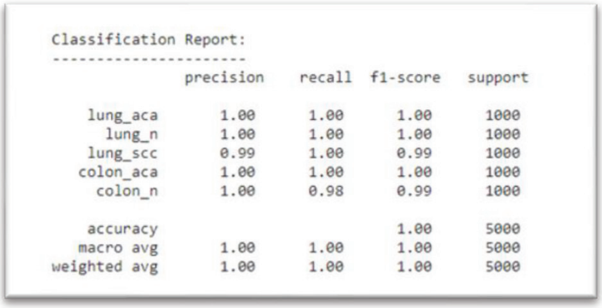
5. Results of the final model

Once the best model was chosen from the training made on our model, the next step was to set a Classification Report (Figure 12) as well as a Confusion Matrix (Figure 13) to measure its performance, showing on the center the accurate predictions it made on the dataset.

6. Conclusions

In this research, several methods for improving the model's classification performance were proposed, testing them and proving their effectiveness on the selected datasets. The result of this study saves exploration time when creating new models for cancer detection, as well as getting better performance in practical applications in real life.

The purpose of this research was to test various optimizers namely Adagrad, Adam, SGD + Momentum, and SGD + Nesterov.



Classification Report:
----

	precision	recall	f1-score	support
lung_aca	1.00	1.00	1.00	1000
lung_n	1.00	1.00	1.00	1000
lung_scc	0.99	1.00	0.99	1000
colon_aca	1.00	1.00	1.00	1000
colon_n	1.00	0.98	0.99	1000
accuracy			1.00	5000
macro avg	1.00	1.00	1.00	5000
weighted avg	1.00	1.00	1.00	5000

Figure 12. Classification Report of Model Performance

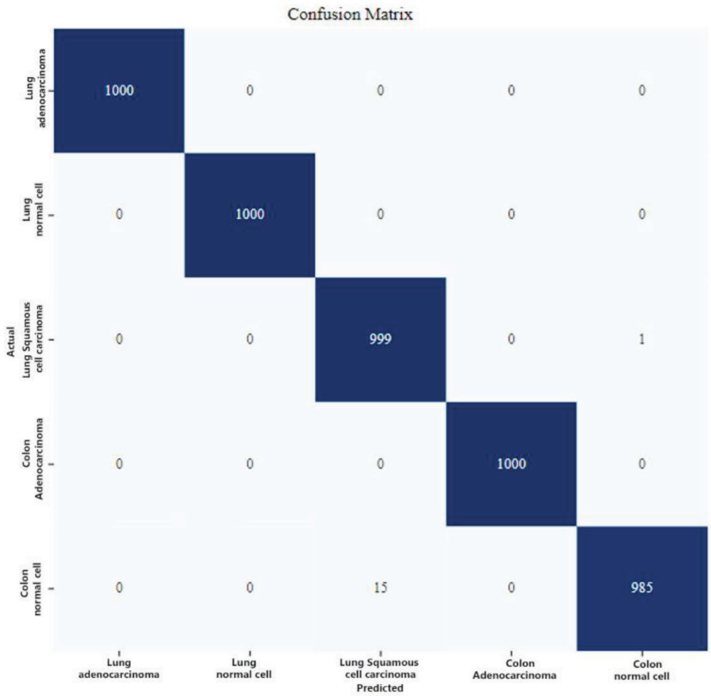


Figure 13. Confusion Matrix from final model on selected optimizer (Adam)

While Adam, SGD + Momentum, and SGD + Nesterov optimizers took their time to achieve good values for Accuracy and Loss, Adagrad obtained good results on early epochs but reached stagnation during some sessions, which could be problematic when working with more classes on larger training periods.

'SGD+Nesterov' showed really good results, yet on scarce occasions, as with 'SGD+Momentum' which also achieved great results, except for those on Kather Dataset (Figure 5).

Finally, and after these experiments, Adam optimizer was selected for having the best performance during the training and validation process, being the one which obtained the best results during this cancer-classification task as it was explained in the Network Training and Result Analysis section.

From the Confusion Matrix (Figure 13), it becomes evident why Adam proved to be the right choice, as the diagonal values of this Matrix show a high rate of accurate prediction hits on each of these 5 classes achieving a perfect score of 1000 right guesses from 1000 samples for 3 of the selected classes(Lung Adenocarcinoma, Lung Normal cell, and Colon Adenocarcinoma).

On the rows arranged next to the principal diagonal of this Matrix, are the wrong predictions which did not match with the actual class, yet remaining competitive results with 999 right guesses on Lung Squamous Cell Carcinoma class, and 985 correct predictions from Lung Squamous Cell Carcinoma class, this last one being the lowest between the 5 classes.

The use of Deep Learning on medical fields such as Cancer recognition models have many challenges; since every cell is different and cancer works randomly on cells, the use of techniques to improve the input data allows to overcome these obstacles, reaching higher accuracy values as well as a better generalization ability. Still, the algorithm proposed on the current study has shown promising features, such as high precision and low loss rate, runs relatively fast, combining different techniques and approaches as a deep learning algorithm.

With the aim of solving the problems of low detection accuracy of cancerous targets, the use of Depthwise Separable Convolutions was selected for this CNN model, since its architecture has fewer parameters than those on commonly used, thus is less prone to overfitting.

Although the accuracy of cancer detection has improved to a certain extent, further studies must be carried out to improve and perfect these models and techniques to achieve better and optimal results on this task. The success of this research motivates future in-depth exploration of the dynamics, with different optimizer and generalization in later epochs where the optimal convergence is harder to reach.

References

Alwosheel, A., van Cranenburgh, S., and Chorus, C. G., 2018. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. Journal of Choice Modelling, 28, 167–182. doi:10.1016/j.jocm.2018.07.002.

Alzubaidi, L., Zhang, J., Humaidi, A.J., et al., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53. https://doi.org/10.1186/s40537-021-00444-8.

Bengio, Y., Courville, A., and Vincent, P., 2014. Representation Learning: A Review and New Perspectives. [cs.LG]. Opgehaal van http://arxiv.org/abs/1206.5538.

Bengio, Y., 2012. Practical Recommendations for Gradient-Based Training of Deep Architectures. Neural Networks: Tricks of the Trade.

Brownlee, J., 2021. Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. Machine Learning Mastery. Retrieved February 7, 2022, from https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/.

Borkowski A. A., Bui M. M., Thomas L. B., Wilson C. P., DeLand L. A., and Mastorides S. M., 2019. Lung and Colon Cancer Histopathological Image Dataset (LC25000). https://doi.org/10.48550/arXiv.1912.12142 [eess.IV].

Candemir, S., Nguyen, X. V., Folio, L. R., and Prevedello, L. M., 2021. Training Strategies for Radiology Deep Learning Models in Data-limited Scenarios. Radiology: Artificial Intelligence, 3(6), e210014. https://doi.org/10.1148/ryai.2021210014.

Chan, H. P., Hadjiiski, L. M., and Samala, R. K., 2020. Computer-aided diagnosis in the era of deep learning. Medical physics, 47(5), e218–e227. https://doi.org/10.1002/mp.13764) (Sarah J. MacEachern and Nils D. Forkert. Machine learning for precision medicine. Genome. 64(4): 416-425. https://doi.org/10.1139/gen-2020-0131.

Cheng, J., Benjamin, A., Lansdell, B., and Kordin, K. P., 2021. Augmenting Supervised Learning by Meta-learning Unsupervised Local Rules. CoRR, abs/2103.10252. Opgehaal van https://arxiv.org/abs/2103.10252.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Dönicke, T., Lux, F., and Damaschk, M., 2019. Multiclass Text Classification on Unbalanced, Sparse and Noisy Data.

Elashmawi, H., 2019 «Optimization of Mathematical Functions Using Gradient Descent Based Algorithms». Mathematics Theses. 4. https://opus.govst.edu/theses_math/4.

Gylberth, R., 2018. Momentum Method and Nesterov Accelerated Gradient - Konvergen.AI. Medium. Retrieved February 6, 2022, from https://medium.com/konvergen/momentum-method-and-nesterov-accelerated-gradient-487ba776c987.

Hirsch, F. R., Franklin, W. A., Gazdar, A. F., and Bunn, P. A., 2001. Early Detection of Lung Cancer: Clinical Perspectives of Recent Advances in Biology and Radiology. Clinical Cancer Research, 7(1), 5–22. Opgehaal van https://clincancerres.aacrjournals.org/content/7/1/5.

Hussain Z., Gimenez F., Yi D., and Rubin D., 2017. Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. AMIA. Annual Symposium proceedings. AMIA Symposium. 2017:979-984. PMID: 29854165; PMCID: PMC5977656.

Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., and Snowdon, J. L., 2021. Precision Medicine, AI, and the Future of Personalized Health Care. Clinical and translational science, 14(1), 86–93. https://doi.org/10.1111/cts.12884.

Karpathy, A., 2017. A Peek at Trends in Machine Learning. https://karpathy.medium.com/a-peek-at-trends-in-machine-learning-ab8a1085a106. [Online; accessed 12-Dec-2017].

Kather, J. N., Halama, N., and Marx, A., 2018. 100,000 histological images of human colorectal cancer and healthy tissue (v0.1).

Kather J. N., Weis C. A. , Bianconi F., Melchers S. M., Schad L. R., Gaiser T., Marx A., and Zollner F., 2016. Multi-class texture analysis in colorectal cancer histology. Scientific Reports (in press).

Keskar, N., and Socher, R., 2017. Improving Generalization Performance by Switching from Adam to SGD. https://doi.org/10.48550/arXiv.1712.07628.

Kingma, D. P., and Ba, J., 2015. Adam: A Method for Stochastic Optimization. CoRR. https://doi.org/10.48550/arXiv.1412.6980

Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Reds), Advances in Neural Information Processing Systems (Vol 25). Opgehaal van https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Liu, Z., Xu, Z., Rajaa, S., Madadi, M., Junior, J. C. S. J., Escalera, S., Pavao, A., Treguer, S., Tu, W., and Guyon, I., 2020. Towards Automated Deep Learning: Analysis of the AutoDL challenge series 2019. Proceedings of the NeurIPS 2019 Competition and Demonstration Track, in Proceedings of Machine Learning Research. Available from https://proceedings.mlr.press/v123/liu20a.html.

More, A., 2016. Survey of resampling techniques for improving classification performance in unbalanced datasets. [stat.AP]. Opgehaal van http://arxiv.org/abs/1608.06048.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., et al., 2015. Deep learning applications and challenges in big data analytics. Journal of Big Data 2, 1. https://doi.org/10.1186/s40537-014-0007-7.

Ozawa, T., Ishihara, S., Fujishiro, M., Kumagai, Y., Shichijo, S., and Tada, T., 2020. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therapeutic advances in gastroenterology, 13, 1756284820910659. https://doi.org/10.1177/1756284820910659.

Ruder, S., 2016. An overview of gradient descent optimization algorithms, pp. 11. https://doi.org/10.48550/arXiv.1609.04747

Schmidt, D., 2018. Understanding Nesterov Momentum (NAG). https://dominikschmidt.xyz/nesterov-momentum/

Scholte, M., van Dulmen, S. A., Neeleman-Van der Steen, C. W. M., et al., 2016. Data extraction from electronic health records (EHRs) for quality measurement of the physical therapy process: comparison between EHR data and survey data. BMC Med Inform Decis Mak 16, 141. https://doi.org/10.1186/s12911-016-0382-4.

Shorten, C., and Khoshgoftaar, T. M., 2019. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0

Sutskever, I., Martens, J., Dahl, G., and Hinton, G., 2013. On the importance of initialization and momentum in deep learning. Proceedings of the 30th International Conference on Machine Learning, in PMLR- 28(3):1139-1147.

Taghiakbari, M., Mori, Y., and von Renteln, D., 2021. Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. World journal of gastroenterology, 27(47), 8103–8122. https://doi.org/10.3748/wjg.v27.i47.8103.

Tom, J., Fei-Fei, L., Ranjay, K., Leila, A., Amil, K., and Chen, C. K., 2020. CS231n: Convolutional Neural Networks for Visual Recognition. https://cs231n.github.io/neural-networks-3/.2020.

Wilson, A., Roelofs, R., Stern, M., Srebro, N., and Recht, B., 2017. The Marginal Value of Adaptive Gradient Methods in Machine Learning. NIPS.

Xue Y., Chen S., Qin J., Liu Y., Huang B., and Chen H., 2017. Application of Deep Learning in Automated Analysis of Molecular Images in Cancer: A Survey. Contrast Media & Molecular Imaging, vol. 2017, Article ID 9512370, 10. https://doi.org/10.1155/2017/9512370.

Ye, G. B., and Xie, X., 2012. Learning sparse gradients for variable selection and dimension reduction. Mach Learn 87, 303–355. https://doi.org/10.1007/s10994-012-5284-9.

Zhou, B. C., Han, C. Y., and Guo, T. D., 2021. Convergence of Stochastic Gradient Descent in Deep Neural Network. Acta Math. Appl. Sin. Engl. Ser. 37, 126–136. https://doi.org/10.1007/s10255-021-0991-2.

Zhou, P., Feng, J., Ma., C., Xiong, C., Hoi, S., and Weinan, E., 2020. Towards Theoretically Understanding Why SGD Generalizes Better Than ADAM in Deep Learning. https://doi.org/10.48550/arXiv.2010.05627.

Authors's Biography

	David Zambrano, received a Master Degree in Computer Science and Technology from Harbin University of Science and Technology, located in Harbin, China, in 2022. Following my vocation and Life Project, been working with technology on world of education for 7 years, and research about technological advances such as Deep Learning Besides technology, my interests reside on languages, currently learning Chinese as my fifth language, focusing on taking advantage of languages, such as for obtaining information all around the world, for researching and creating solutions on benefit of society.
	Sun Bowen, University of Science and Technology and a master tutor in computer science and technology. In 1999, he founded the «Fractal Channel» website. Two sets of «Internet Times» programs of CCTV were broadcasted on February 27, 2002 to introduce the «Fractal Channel» website. Sun Bowen 's fractal art works have been exhibited in Tianjin Science and Technology Museum, Heilongjiang Science and Technology Museum and other units many times. Research direction: fractal graphics, the complexity of financial markets.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 285-307
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.27885

FBCHS: Fuzzy Based Cluster Head Selection Protocol to Enhance Network Lifetime of WSN

Vipul Narayan[a] and A. K. Daniel[b]

[a, b] Department of Computer Science and Engineering, Madan Mohan Malaviya University, Gorakhpur, UP, India, 273008

✉ vipulupsainian2470@gmail.com, danielak@rediffmail.com

KEYWORDS

clustering; cluster head; energy efficient; fuzzy logic; residual energy; sensor network

ABSTRACT

With enormous evolution in Microelectronics, Wireless Sensor Networks (WSNs) have played a vital role in every aspect of daily life. Technological advancement has led to new ways of thinking and of developing infrastructure for sensing, monitoring, and computational tasks. The sensor network constitutes multiple sensor nodes for monitoring, tracking, and surveillance of remote objects in the network area. Battery replacement and recharging are almost impossible; therefore, the aim is to develop an efficient routing protocol for the sensor network. The Fuzzy Based Cluster Head Selection (FBCHS) protocol is proposed, which partitions the network into several regions based on node energy levels. The proposed protocol uses an artificial intelligence technique to select the Cluster Head (CH) based on maximum node Residual Energy (RE) and minimum distance. The transmission of data to the Base Station (BS) is accomplished via static clustering and the hybrid routing technique. The simulation results of the FBCHS protocol are com- pared to the SEP protocol and show improvement in the stability period and improved overall performance of the network.

1. Introduction

In the last few years, designing an efficient network has become important in cloud computing, IoT, auto- mobile, precision agriculture, artificial intelligence, etc. The WSN consists of multiple Sensor Nodes (SNs) and wireless technology. The distribution of nodes is random for the data collection task in the network. The data is collected from sensors which measure a series of environmental parameters: temperature, hu- midity, rain, fog, and transmitted to BS. In Narayan and Daniel, 2019, the authors said that communication via single-hop sends data directly while intermediate nodes follow multi-hop communication. The sensing and data transmission operations consume significant energy, which causes energy issues in the network. In Narayan and Daniel, 2020, the authors said that the hierarchical architecture minimizes the problems. Clustering is the technique that divides the network into smaller cluster groups to increase system performance. The CH is selected and plays an essential role in processing, communication, and data transmission tasks in the network. In Deepa and Suguna, 2020, the authors said that clustering is effective communication strategies and in minimizing the energy issues in the network. It also helps prevent frequent collisions and overhead and controls data congestion in the network. In Narayan and Daniel, 2021, it is stated that in the existing clustering techniques, the CH was selected on the basis of RE of Sensor Node (SN). Therefore, an energy-efficient solution was introduced to limit the network's energy consumption for CH nodes. In Elhoseny et al., 2020, the authors said that various CH selection approaches were utilized to reduce network energy issues and increase system performance. The CH was chosen randomly among the nodes, adding density as one more criterion and transmitting the data to BS. There is a need to develop an efficient routing protocol that utilizes minimum energy dissipation of nodes inside the network, and that reliably forwards the data to the BS.

The proposed FBCHS protocol is introduced. The static clustering and hybrid routing mechanism is used to transmit data to BS. The Normal Node (N) and Super Node (S) are dispersed in the different regions depending on the SNs energy level for proficient coverage of the total area of the network. The hybrid routing technique is utilized for the purposes of data transmission to the BS, where the Normal Nodes communicate directly to the BS. The clustering technique transmits data to BS in Super Nodes of Region- 1 (R1) to Region-4 (R4). On the basis of the Artificial Intelligence (AI) approach, the fuzzy logic technique is applied for the CH selection process. The SN uses parameters such as RE of node and minimum node distance to the BS. Thus, AI-based CH selection balances network energy consumption inside the network and enhances network longevity. The performance of the FBCHS protocol is contrasted with the SEP protocol in terms of the stability period, network performance, alive nodes number, a packet transmitted to BS, percentage of the covered area, and network data transmission per round. The remainder of the paper is organized as follows: Section 2 discusses the issues associated with design and related work. The radio model is presented in Section 3. Section 4 proposes the Fuzzy Based Cluster Head Selection Protocol. Section 5 outlines simulation and analytical results, and finally concludes the paper with future lines of research.

2. Related Work

The WSN comprises several nodes that operate in a specific environment for data collection and transmission via single or multi-hop techniques in the network. The data transfer from CH to BS is an im- portant task in a sensor network. Various clustering techniques have been proposed in the last decades for WSN. The Low Energy Adaptive Clustering Hierarchy (LEACH) protocol uses certain probability parameters for CH selection per round. Each node has an equal probability of selecting a CH in the protocol and follows a single hop for data transmission to BS. In Nazari Talooki et al., 2014, the authors proposed a clustering technique for the uniform distribution of energy across the network. The technique has a self-organized feature and minimizes load balancing problems in multi-hop routing techniques. The pioneer clustering scheme is in the network. The selection of the CH is performed using the probabilistic technique. The multitier communication architecture is used for data transmission tasks in the network. In the first step, cluster formation takes place and relays data to their cluster leader. In the second step, the CHs aggregate all data and transmit it to BS. CHs selection is not uniform in the network due to its probabilistic nature. Thus, some regions have more CH than others, which causes load balancing problems and affects network performance. In Kumar et al., 2021, the authors proposed the Improved Density Control and Divide Rule (IDDR) protocol. The network area is divided on the basis of node density. The divide and rule techniques were used for efficient energy utilization in the network. Each node's energy and status information were managed by BS. The CH selection process was random in the protocol, and BS' responsibility was to discard low-energy SNs for CH selection. In some cases, BS could not find the node status located far from it. The protocol improved the network's performance compared to the LEACH Protocol. In Latif et al., 2015, the authors introduced a protocol that partitioned the network into regions and a clustering mechanism for efficient data transmission to BS. The network was partitioned into multiple rectangular regions and sub-regions to balance energy consumption. The corner node in the sub-regions experienced a long distance for data transmission to BS than a node nearer to BS. In Singh and Daniel, 2015, the authors proposed a technique to address the issue of load balancing in sensor networks. SNs were randomly distributed in the multiple square regions, and static clustering was performed for the heterogeneous nodes. The protocol selected CH depending on the RE of SNs. The protocol enhanced the network lifetime but did not solve the larger problem of network holes. The various fault tolerance routing protocols were discussed. Furthermore, the majority of known fault tolerance techniques include hardware and path redundancy in the network. The Dual-Homed fault-tolerant Routing (DHR) protocol was proposed to solve the problem in the network. The dual CH selection mechanism was used and enhanced the performance of the network system. In Latif et al., 2016, in a large network area, the authors proposed a routing strategy to reduce node energy usage inside the network. The protocol self-adjusts the node's power for a small distant node within the area and improves the system's performance. In Yang et al., 2018, the authors proposed a clustering technique in which the network was partitioned into unequal regions and performs static clustering. The sub-regions were further partitioned in quadrilaterals to minimize the distance within the network. The load balancing was achieved, however, there were overhead increases in the protocol. In Behzad, 2018, the authors proposed a protocol for the analysis of the network lifetime based on the energy models. It was observed that theoretical analysis has more energy values in different network areas. In Narayan and Daniel, 2022, the authors proposed an algorithm for optimal energy inside the network. The working protocol was the same as LEACH but used a multi- hop technique for data transfer in the network. The protocol reduced the communication distance between CHs and improved network performance. In Sharma and Kumar, 2018, the authors proposed an energy-based routing protocol to reduce cluster node transmission distance. The new energy model is used in which the transmission distance is varied by changing the cluster position in the protocol and enhancing the system's performance. In Sirsikar and Chandak, 2018, the authors proposed a model in which the network was partitioned in the rectangular region and concentric ring. The static clustering technique was used to transmit data to BS. The simulation performance of the partitioned rectangular region was better in comparison to concentric ring in terms of lifetime. In Tang et al., 2018, the authors proposed Physarum-based protocol using the energy harvesting technique. The protocol used a distributed routing algorithm to reduce time and minimize energy usage in the network. The hot spot problem was minimized by using a distributed algorithm and unequal clustering. The Fuzzy-AI technique simulated the problem in a more precise manner, solving service selection problems in the network. In Faiz and Daniel, 2021, the authors proposed the Fuzzy model. The model selected the user-defined input parameters on the basis of selected parameters; the best cloud offered to users for service. The Fuzzy concept was used for the CH selection for efficient energy utilization among SNs in WSN applications. In Jayaraman and Dhulipala, 2021, the authors proposed a Fuzzy Logic-based CH selection mechanism. CH energy, the centrality of nodes, the distance between nodes, and the base station were used as fuzzy variables because of their Mamdani fuzzy inference system. The mechanism utilized optimal clusters and selected CH on the basis of the maximum RE of the node in the network. The protocol achieved load balancing, efficient utilization of energy inside the network, and enhanced network performance. In Batra and Kushwah, 2019, the authors proposed a fuzzy logic based CH selection method to extend the network lifetime. The protocol utilized optimal nodes, shortest distance between CH and BS, maximum balanced energy after each round, and CH rotation in the network. The k-means method was utilized for cluster formation in the network, and the CH was selected using fuzzy logic to increase system performance. In Shivappa and Manvi, 2019, the authors proposed a centralized CH selection mechanism and fuzzy approach for forming clusters in the network. The BS node used fuzzy c-means to find cluster centers and member nodes. The fuzzy FIS calculated the nodes' fitness, RE of SNs and distance to the cluster center as input parameters. The CH is chosen from each cluster by using distance and maximum node RE as parameters. The proposed mechanism utilized optimal nodes inside the network and improved system performance. In Choudhary et al., 2022, the authors proposed fuzzy base routing protocol for the sensor network. The quality-of-service parameters for sensor network was enhanced by utilizing most trusted values consideration for CH selection. The CH selection was based on the fuzzy if/then rule for optimal route selection for data transmission to BS and significantly improved system performance. In Dwivedi and Sharma, 2020, the authors have proposed a fuzzy base protocol for efficient clustering and CH selection in the network. The fuzzy-based rules were used for optimal cluster inside the network and triangular inequality theorem for shortest distance for data transmission between target and base station. The proposed protocol minimized the transmission distance and enhanced system performance.

3. Radio Model

The radio model transmits and receives the 'L' data bits in WSN. In terms of how far away the transmitter and receiver are from each other, the radio model can be called a free space or multipath model. Figure 1 represents the radio model (Wan and Du, 2011; Mehra, Doja and Alam, 2020).

If the threshold distance 'd_0' is more than the distance 'd', then the transmission energy required for the free space model is:

$$E_{TX} = L \times E_{elec} + L \times \varepsilon_{fs} \times d^2 \; d << d_0 \quad (1)$$

If the distance '$d_{0'}$ is less than the threshold distance 'd,' then the transmission energy required for the multipath model is:

$$E_{TX} = L \times E_{elec} + L \times \varepsilon_{mp} \times d^4 \; d >> d_0 \quad (2)$$

Where,

L is the bits needed for sending data from one sensor to another, ε_{fs} & ε_{mp} represent the amount of power for free space and multipath models, and E_{elec} is the electronic energy required for the signal and

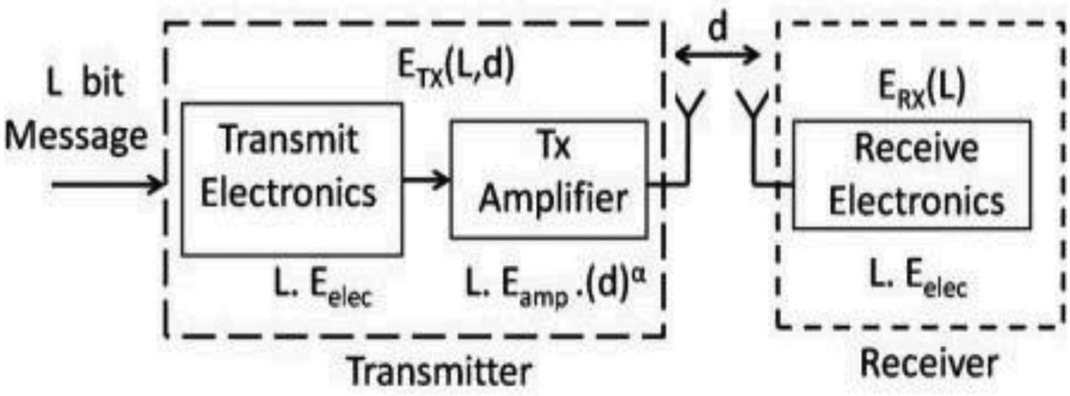


Figure 1. The Radio Model

$$d_0 = \sqrt{\varepsilon_{fs} / \varepsilon_{mp}}$$

The received energy is calculated as follows:

$$E_{RX} = L \times E_{elec} \tag{3}$$

The CM energy is:

$$E_{CM} = L \times E_{elec} + L \times \varepsilon_{fs} \times CH_d \tag{4}$$

The CH energy is:

$$E_{CM} = L \times count \times (E_{elec} + L \times \varepsilon_{fs} \times BS_d + E_{AE} \tag{5}$$

CH_d is the distance from CH, BS_d is the distance from BS, and E_{AE} is the total aggregated energy.

4. Proposed FBCHS protocol

The data gathering process is efficiently performed when the coverage area is large. If the monitoring area is large, some portion of the area is uncovered and called a coverage hole. The one possible solution to this issue is to partition the network into multiple regions on the basis of their RE levels of SNs and fixed coordinates of the X–Y dimensions. In the process of deploying SNs in the network area, some SNs are deployed far from the area, and the corner node needs high energy to relay the data to the BS. Therefore, to solve this issue, high-energy nodes called Super Nodes (S) are deployed in far distance places, and Normal Nodes (N) are positioned near the BS in Region 5. Let us suppose that SNs are the total nodes in the network area, and the fraction of Super Nodes is S. The Normal Node Initial Energy (IE) is 'E_0', and the Super Node has ($1+E_{\alpha}$) more energy than Normal Node. In Region 5, the Normal Node transmit data to BS in a single hop, while Super Nodes in Regions 1, 2, 3, and 4 send their data through CH. The Normal Node dies quickly because it consumes more energy than indirect data transmission tasks in the network. The direct transmission takes more energy; therefore, the clustering technique is used for SN to transmit data to BS. The FBCHS Protocol uses the optimal number of CH per round for Super Node regions and enhances the network performance.

4.1. Fuzzy Based Cluster Head Selection (FBCHS) protocol

Consider a network with a size of 100 × 100 m^2 and a location of BS in the approximate middle (50 × 50 m^2). The network is partitioned into different regions (Region 1 to Region 5) with fixed coordinates. The SNs are distributed uniformly in the fixed predefined regions. The Normal Node and Super Node are used in the predefined regions. The position of the Super Node and the BS is static following the deployment process. The Normal Node transmits data to BS while the Super Node collects all data from cluster nodes and then transmits it to the BS.

The proposed FBCHS protocol partitioned the network into five regions R_1, R_2, R_3, R_4, R_5 as shown in Figure 2. The SNs are randomly distributed in the network of 100 × 100 m^2. The used SNs are Super Nodes and Normal Nodes. Regions R_1, R_2, R_3, and R_4 comprise Super Nodes, whereas region R_5 comprises Normal Nodes. Thus, the Normal and Super Nodes are divided in the ratio of 7:3 in the entire network. The regions R_1, R_2, R_3, R_4, R_5 divide the network in the ratio of 20:20:12:12:36. The relationship of regions with coordinates a, b, c, and d is represented below.

$$R_1: x \in (a, b), y \in (a, d)$$
$$R_2: x \in (c, d), y \in (a, d)$$
$$R_3: x \in (b, c), y \in (a, b)$$
$$R_4: x \in (b, c), y \in (c, d)$$
$$R_5: x \in (b, c), y \in (b, c)$$

Where, coordinates a = 0, b = 20, c = 80, d = 100.

4.2. CH Selection Formula

The Probability of Normal (P_N) and Super (P_S) Nodes to be CH are, respectively.

$$P_N = \frac{P_O}{1+\Phi\times\Psi} \tag{6}$$

$$P_S = \frac{P_O}{1+\Phi\times\Psi}\times(1+\Phi) \tag{7}$$

Where φ stands for the additional fraction of energy present in Super Node (as opposed to the Normal Node), T is the fraction of SNs over the total SNs. The thresholds for Normal SNs (T_N) and Super SNs (T_S) are defined as:

$$T_N = \begin{cases} \dfrac{P_N}{1-P_N(r \bmod \frac{1}{P_N})}\times\dfrac{R_N}{D_N}\times\delta' & , N(i)\in G' \\ 0 & , otherwise \end{cases} \tag{8}$$

$$T_S = \begin{cases} \dfrac{P_S}{1-P_S(r \bmod \frac{1}{P_S})}\times\dfrac{R_N}{D_N}\times\delta' & , N(i)\in G'' \\ 0 & , otherwise \end{cases} \tag{9}$$

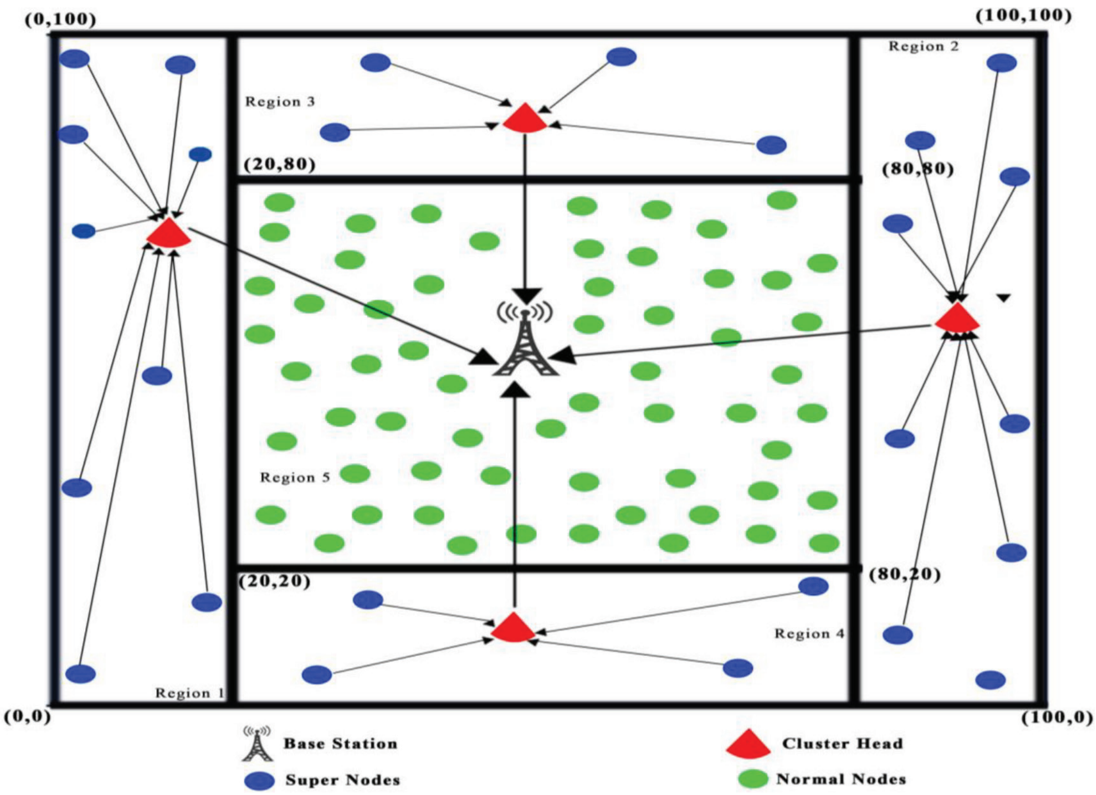


Figure 2. Network Architecture of Fuzzy Based Cluster Head Selection (FBCHS) protocol

Where, R_N is Residual Energy, D_N is the Distance from BS. D_N and δ' are expressed as:

$$D_N = \sqrt{(N_x - BS_x)^2 + (N_y - BS_y)^2} \tag{10}$$

$$\delta' = \sqrt{\frac{N_i}{2\Pi}} \sqrt{\frac{E_{fm}}{E_{mm} \times D_N^4 \times (2\omega - 1) \times E_{T,R} - \omega \times E_D}} \times \beta \tag{11}$$

The notations are defined in following Table 1.

4.3. Fuzzy based CH Selection using AI-Technique

To solve the load balance problem in the WSN, a fuzzy-based technology called «Fuzzy-AI» is introduced. CH selection in WSN routing protocols is based on a single criterion, however, we have taken two parameters; RE of SNs and minimum BS distance for the FBCHS protocol. The input function

Table 1. Notations and Descriptions

Symbols	Meaning
P_O	Optimal Probability for Number of Cluster Heads.
Φ	Additional Fraction of Energy in Super Node.
T	Fraction of SNs Over the Total SNs.
P_N	Probability for Normal Node to be CH.
P_S	Probability for Super Node to be CH.
T_N	The Threshold for Normal Node.
T_S	The Threshold for Super Node.
N_x	x Coordinates of i^{th} Node.
N_y	y Coordinates of i^{th} Node.
BS_x	x Coordinates of Base Station.
BS_y	x Coordinates of Base Station.
D_N	i^{th} Node Distance from BS.
R_N	Residual Energy.
δ'	Optimal Cluster Number.
N_i	i^{th} Node.
E_{mf}	Free Space Energy.
E_{mm}	Multipath Fading Energy.
ω	The Number of Associated Cluster Heads.
$E_{T,R}$	Energy Transmission/ Reception.
E_D	Data Aggregation Energy.
β	Network Size.

chosen for CH selection is the distance to BS and the RE of the SNs, and they are converted into their respective fuzzy sets. The mathe- matical representation of the same is as follows:

$$\text{Distance} = \{(dt, \mu_{\text{Distance}}(dt))\ d \in D \tag{12}$$

$$RE = \{(rl, \mu_{RE}(rl))\}\ rl \in RE \tag{13}$$

Where D and RE are the universes of discourse for Distance and node Residual Energy, respectively. Dt and rl represent the individual element of sets 'D' and 'RE', $\mu_{\text{Distance}}(dt)$ and $\mu_{RE}(rl)$ are membership functions. The process of CH selection takes place after the successful deployment of SNs in the monitoring area.

The fuzzy logic model is utilized to solve uncertainty in CH selection. The CH is selected using the parameters minimum BS distance, and the best node was chosen to serve as CH (Maurya and Daniel, 2015). Generally, a Fuzzy system consists of different units:

Fuzzifier: In this unit, input variable distance and RE are given crisp input, converted into corresponding fuzzy input using a fuzzy set.

Rule base: The FIS unit stores all IF-THEN-ELSE rules determined by a person with domain expertise. Fuzzy rules represent the active role of the fuzzy system.

Fuzzy Inference Engine (FIS): The key component of a fuzzy system provides a decision-making pro- cess. This unit uses IF-THEN-ELSE rules to design the required decision rules using «AND» or «OR» opera- tor. The FIS for the proposed work is shown in Figure 3.

Defuzzification: In this unit fuzzy set is converted into a crisp set with the help of FIS and its rule base.

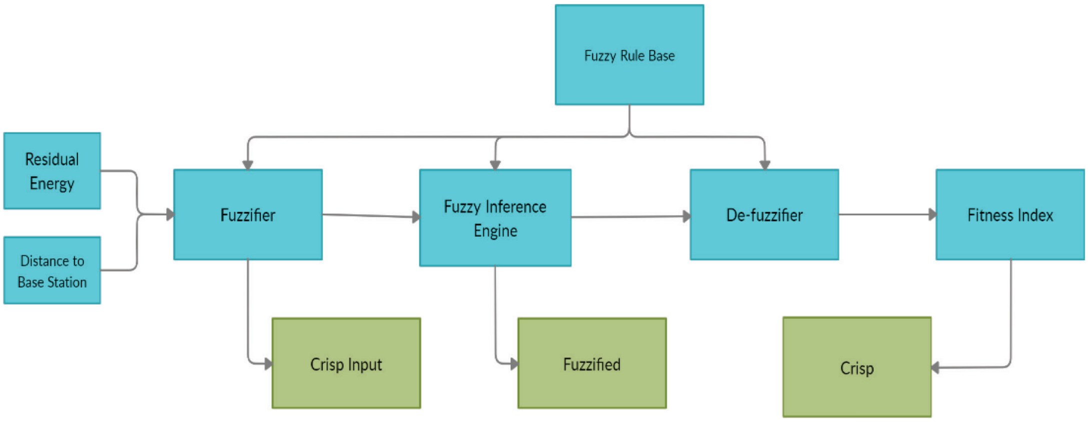


Figure 3. Block Diagram of Fuzzy Inference Engine (FIS)

A. *Fuzzification of Input*

This phase converts crisp values into fuzzy variables. Table 2 shows the crisp input range values for 'Dis- tance' on a scale of 0-100, where the Fuzzy output variable represents the membership values Short, Consid- erable, and Far as shown in Figure 4.

Table 2. Range for Fuzzy input parameters (Distance)

S.No.	Crisp Value	Fuzzy variable
1	[0 0 45]	SHORT
2	[40 55 65]	CONSIDERABLE
3	[60 85 100]	FAR

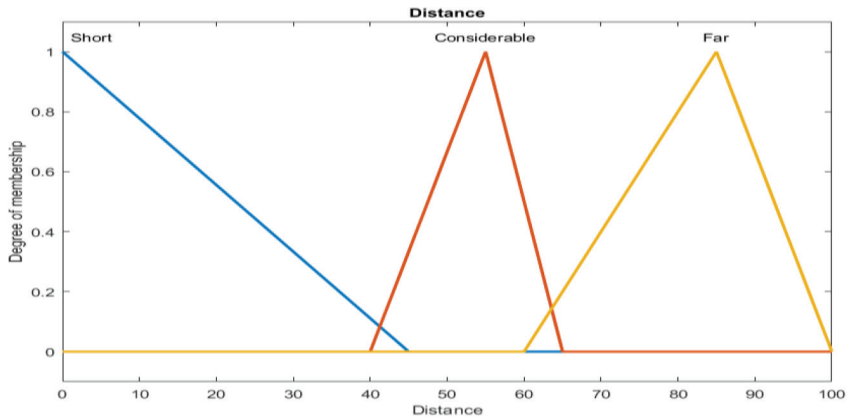


Figure 4. Membership function for 'Distance' on the scale of 0-100

Similarly, Table 3 shows the crisp input values for 'Residual Energy' on a scale of 0– 1. The fuzzy output variable shows the membership value Low, Reasonable, and High as shown in Figure 5.

Table 3. Range for Fuzzy input parameters (RE)

Sr. No.	Crisp Value	Fuzzy variable
1	[0 0 0.3]	LOW
2	[0.25 0.6 0.8]	MEDIUM
3	[0.75 0.85 1]	HIGH

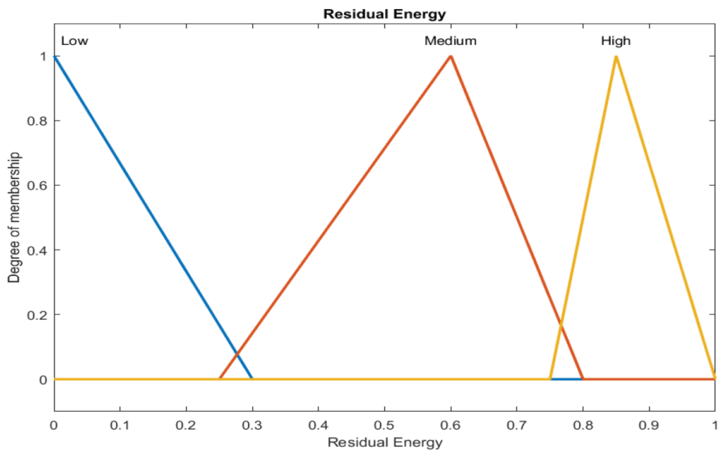


Figure 5. Membership function for 'RE' on the scale of 0-1

B. *Output Membership Function*

The membership functions are μA(Dt), and μB(RE). Both variables are associated with Lower and Upper threshold values. The respective thresholds are TH_1, TH_2, TH_3 and TH_4. The activation point of the model is a lower threshold value, and the model operates in both the upper and lower threshold regions. [14].

$$\mu_x(Distance) = \begin{Bmatrix} 1 & \text{if } Dt \leq TH_1 \\ TH_1 - Dt \big/ TH_2 - TH_1 & \text{if } TH_1 < Ds < TH_2 \\ 0 & \text{if } Dt \geq TH_2 \end{Bmatrix} \quad (14)$$

$$\mu_y(R.E) = \begin{Bmatrix} 0 & \text{if } RE \leq TH_3 \\ RE - TH_3 \big/ TH_4 - TH_3 & \text{if } TH_3 < RE < TH_4 \\ 1 & \text{if } RE \geq TH_4 \end{Bmatrix} \quad (15)$$

Let Dt and RE be represented for the inputs as Distance (Dt) and Residual Energy (RE), respectively. The fuzzy variables for output representation, using the Membership Function (MF) for CH selection, are used. 'VLOW' represents very low membership, 'LOW', 'MEDIUM', 'HIGH' and 'VHIGH' represents very high membership, as shown in Table 4. Figure 6. shows Fuzzy MF variables for CH selection on a scale of 0-10. The FBCHS system for computing likelihood is shown in Figure 7.

Table 4. Fuzzy MF variables for CH

No.	Linguistic Term	Values
1	[0 0 1 2]	VLow
2	[1.5 2.5 3 3.5]	Low
3	[3.2 4.5 5 5.5]	Medium
4	[5.2 6 6.5 7]	High
5	[6.8 7.5 9 10]	VHigh

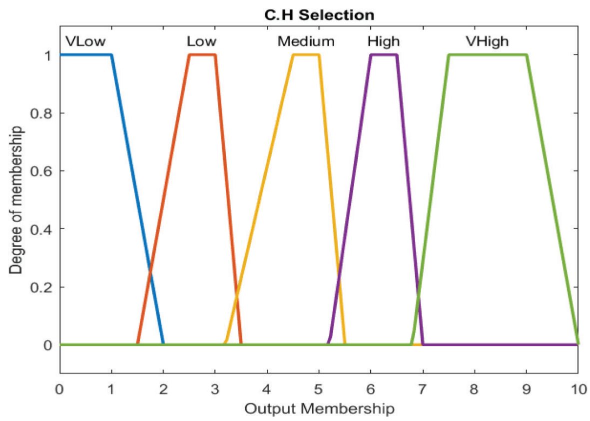


Figure 6. Membership function for 'CH Selection' on the scale of 0-10

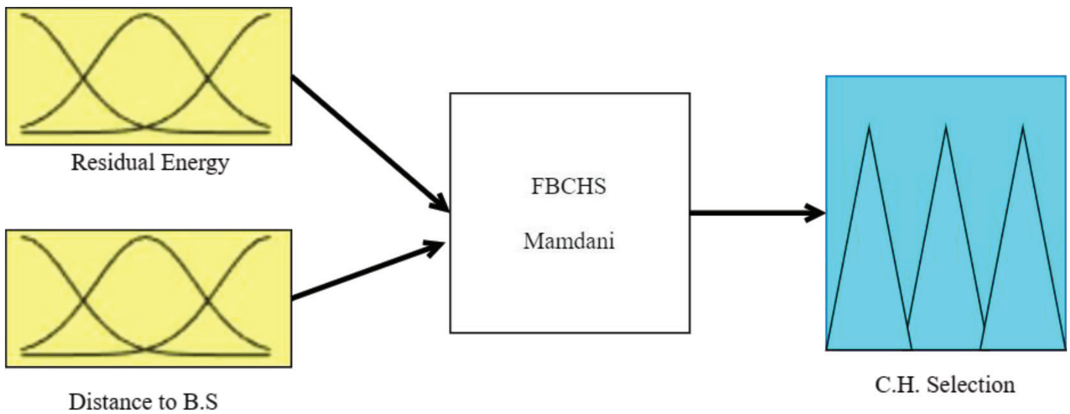


Figure 7. FBCHS system for computing the likelihood

C. Evaluation of Rules for CH Selection

The two parameters used for CH selection in Super Nodes are distance to BS and RE of the Super Node. Three membership functions are used to represent different degrees of the input parameters are shown in Table 5. The CH column of Table 5 shows various output functions.

Table 5. Fuzzy rules

Rules	**Distance to BS**	**RE**	**CH**
1	Short	Low	Medium
2	Short	Medium	Medium
3	Short	High	VHigh
4	Far	Low	VLow
5	Far	Medium	Medium
6	Far	High	Medium
7	Considerable	Low	Medium
8	Considerable	Medium	High
9	Considerable	High	High

The MFs are selected using defined Fuzzy rules embedded in the FIS. The order used for the priority of parameters for selecting a CH is defined using equation (16) as follows:

$$Dt > RE \quad (16)$$

4.4. Fuzzy Inference Rules

1. IF Dt is Short Λ RE is Low; CH Selection probability is ‘Medium’.
2. IF Dt is Short Λ RE is Medium; CH Selection probability is ‘High’.
3. IF Dt is Short Λ RE is High; CH Selection probability is ‘VHigh’.
4. IF Dt is Far Λ RE is Low; CH Selection probability is ‘VLow’
5. IF Dt is Far Λ RE is Medium; CH Selection probability is ‘Medium’
6. IF Dt is Far Λ RE is High; CH Selection probability is ‘Medium’
7. IF Dt is Considerable Λ RE is High; CH Selection probability is ‘High’
8. IF Dt is Considerable Λ RE is Medium; CH Selection probability is ‘High’
9. IF Dt is Considerable Λ RE is Low; CH Selection probability is ‘Medium’

Algorithm 1: FBCHS Protocol

Initialization: R = Region, OA = Outer Area, IA = Inner Area, SNs = Sensor Nodes, S = Super Nodes, N = Normal Nodes, RE = Residual Energy, D = Distance, CH = Cluster Head, CM = Cluster Member, BS = Base Station.

```
For all r in R_i do
If (r ∈ OA)
SNs←S
    Calculate RE
    Calculate D
        For all i in SNs_r do
            Select max (RE)
            Select min (D)
                If (max (RE) && min (D))
i← CH
                else
i←CM
                End If
            End For
    else
SNs←N
    End If
End For
For all r in Ri do
    For all i in SNs_r do
        If (i∈N&&i∉ CH)
            Aggregate (data)
            Send (data → BS)
End If
    If (i∈N&&i∉ CH)
        Send (data → CH)
End If
    If (i∈ N)
```

```
        Send (data → BS)
    End If
      End For
    End For
```

4.5. VALIDATION

If a Normal Nodes senses data, it directly transmits the data to the BS while Super Nodes transmit the data using the clustering technique and then transmit it to the BS. The demonstration for CH node selection using five Super Nodes is an example of understanding the whole process.

Let the set represent Super Nodes using distance from the BS as Dt = {45, 55, 65, 85, 100}, and equation (14) is used for the computation of fuzzy membership values as follows:

$$\mu_x(Distance) = \left\{ \begin{array}{cc} 1 & \text{if } \mathrm{Dt} \leq 60 \\ \mathrm{Dt} - Th1 \big/ Th2 - Th1 & \text{if } 60 < Dt < 80 \\ 0 & \text{if } \mathrm{Dt} \geq 80 \end{array} \right\} \tag{17}$$

Where, $Th_1 = 60$, $Th_2 = 80$, and Dt values are measured in meters.

Let the set represent the Residual Energy (RE) of each node as RE = {0.3, 0.25, 0.6, 0.85, 0.95}, and equation (15) is used for the computation of corresponding fuzzy membership values as follows:

$$\mu_y(RE) = \left\{ \begin{array}{cc} 0 & \text{if } \mathrm{RE} \leq 0.3 \\ \mathrm{RE} - Th3 \big/ Th4 - Th3 & \text{if } 0.3 < RE < 0.8 \\ 1 & \text{if } \mathrm{RE} \geq 0.8 \end{array} \right\} \tag{18}$$

Where, $Th_3 = 0.3$, $Th_4 = 0.8$, and RE values are measured in joules.

Now, compute the membership values of Dt and RE with the help of Membership Functions (MF) for the input values shown in Tables 6 and 7. In Table 6, for the Dt {45, 55, 65, 85, 100}, the values of memberships are{1, 1, 0.25, 0, 0}, respectively. Table 7 shows the RE set {0.3, 0.25, 0.6, 0.85, 0.95} and, the memberships vlaues as {0, 0.6, 1, 1, 1}.

Table 6. Membership value of variable 'Dt'

Dt	**Membership value**
45	1
55	1
65	0.25
85	0
100	0

Table 7. Membership of variable 'RE'

RE	Membership value
0.25	0
0.3	0
0.6	0.6
0.85	1
0.95	1

The AND (Λ) operator is used to compute the fuzzy relation. The membership values of Dt and RE of sensor nodes are listed in Table 8.

Table 8. Membership of Dt AND (Λ) RE using Fuzzy process

Dt	45	55	65	85	100
RE					
0.25	0 Λ 1	0 Λ 1	0Λ 0.25	0Λ 0	0Λ0
0.30	0 Λ 1	0 Λ 1	0Λ0.25	0Λ0	0Λ0
0.60	0.60Λ1	0.60Λ1	0.60Λ 0.25	0.60Λ 0	0.60Λ0
0.85	1Λ1	1Λ1	1Λ0.25	1Λ0	1Λ0
0.95	1Λ1	1Λ1	1Λ0.25	1Λ0	1Λ0

The outcome of Λ operation applied to the Membership Values (MV) of Dt and RE is represented in Table 9.

Table 9. Outcome of the fuzzy (Λ) operation on MV

Dt	45	55	65	85	100
RE					
0.25	0	0	0	0	0
0.3	0	0	0	0	0
0.6	0.6	0.6	0.25	0	0
0.85	1	1	0.25	0	0
0.95	1	1	0.25	0	0

Table 10 shows the various combinations of maximum membership input values of Dt and RE.

Table 10. Fuzzy output for RE

Dt	45	55
RE		
0.85	1	1
0.95	1	1

The Membership Function (MF) for Dt and RE as per the defined rules is represented in Table 11.

Table 11. Degree of MF (Fuzzy Output)

Sl. No.	**Dt**	**Membership function (MF)**	**RE**	**Membership function (MF)**
1	45	Short	0.85	High
2	55	Short	0.85	High
3	45	Short	0.95	High
4	55	Short	0.95	High

The total possible arrangements of Distance (Dt) and Residual Energy (RE) of reachable nodes are as follows:

• Dt_1 = 45 and RE_1 = 0.85 • Dt_2= 55 and RE_2 = 0.85 • Dt_3= 45 and RE_3 = 0.95 • Dt_4= 55 and RE_4= 0.95

The node to be elected as CH can be any possible arrangement among the Super Nodes. As per rule 1, using the defined rule set, the best arrangement is possible when the value of Dt to BS is 45, and the RE of the node is = 0.95.

The output membership value for this combinational arrangement is «VHigh» as «Dt is Short and RE is High». This combination determines the best node to be chosen for the election of CH from among super cluster nodes.

5. Simulation Results and Validations

The analytical-based work is discussed in the previous section. The performance of the FBCHS Protocol is simulated in MATLAB. The 100 SNs are distributed randomly in a 100 m^2 network, and the position of BS is equipped in 50×50 m^2. The heterogeneous SNs (i.e., Super Node and Normal Node) are distributed randomly for 20000 rounds of packet transmission in the network. 70% of SNs are Normal Nodes, and the rest are Super Nodes. The numbers of SNs deployed in the different regions are shown in Table 12, and the per-key parameter for the simulation is shown in Table 13. The performance of the FBCHS protocol contrasts with the SEP protocol for heterogeneous environments and 20000 rounds of packet transmission to BS (Narayan and Daniel, 2021). The simulation performance of the FBCHS protocol outperforms better results than the SEP protocol and enhances the network lifetime.

Assumptions of FBCHS Protocol.

1. The proposed protocol is designed using heterogeneous nodes.
2. The noise factor and collision are ignored in the system.
3. The Cluster head(s) aggregate data and transmit it to BS.
4. The SNs are distributed randomly and deterministic.
5. The battery is not chargeable, and BS continuously supplies power.

Table 12: Number of SNs deployed in regions

Types of Nodes	Nodes Number
Normal Nodes	70
Super Nodes	30
R_1: Super Nodes	10
R_2: Super Nodes	10
R_3: Super Nodes	5
R_4: Super Nodes	5
R_5: Normal Nodes	70

Table 13. Key parameter for the simulation

Parameters	Values
SNs	100
Network Area	100 m^2
Free Space Model (E_{fs})	10pJ/bit/m^2
Multipath Model (E_{amp})	0.0013pJ/bit/m^4
Initial Level Battery (E_0)	0.5 J
Energy Factor α	1
Initial Energy of Advanced Nodes	E_0 (1+E_α).
Electronic Circuitry (E_{RX})	50nJ/bit
Data Aggregation (E_{DA})	10nJ/bit

The simulation is performed for 20000 successful packet transmission rounds in the network. The FBCHS protocol has a greater number of alive nodes than the SEP protocol, as shown in Figure 8(a). All nodes in the SEP protocol die after 4725 rounds, whereas, 25 nodes are alive in the FBCHS, and the network lifetime of the FBCHS protocol is more than 10000 rounds for packet transmission in the network, whereas the network lifetime of the SEP protocol is 4725 rounds, as shown in Figure 8(b). The network lifespan of the FBCHS protocol is 2871 rounds for Normal Nodes, while the network lifespan of the SEP is 2271 rounds for Normal Nodes, as shown in Figure 8(c). The FBCHS protocol Super Nodes network lifetime is greater than 10000 rounds; however, the SEP protocol network lifetime is 4725 rounds, as shown in Figure 8(d). The stability period of the FBCHS protocol is significantly better than SEP, which enhance the system performance. The FBCHS protocol transmitted 1.7×10^5 packets to BS, whereas the SEP protocol transmitted 3×10^4 packets to BS, significantly less than the SEP, as shown in Figure 8(e). The FBCHS protocol covered 100 % of the network for 4100 rounds, whereas the SEP protocol covered only 2552 rounds. The overall FBCHS protocol covered more network area compared to the SEP protocol, as shown in Figure 8(f).

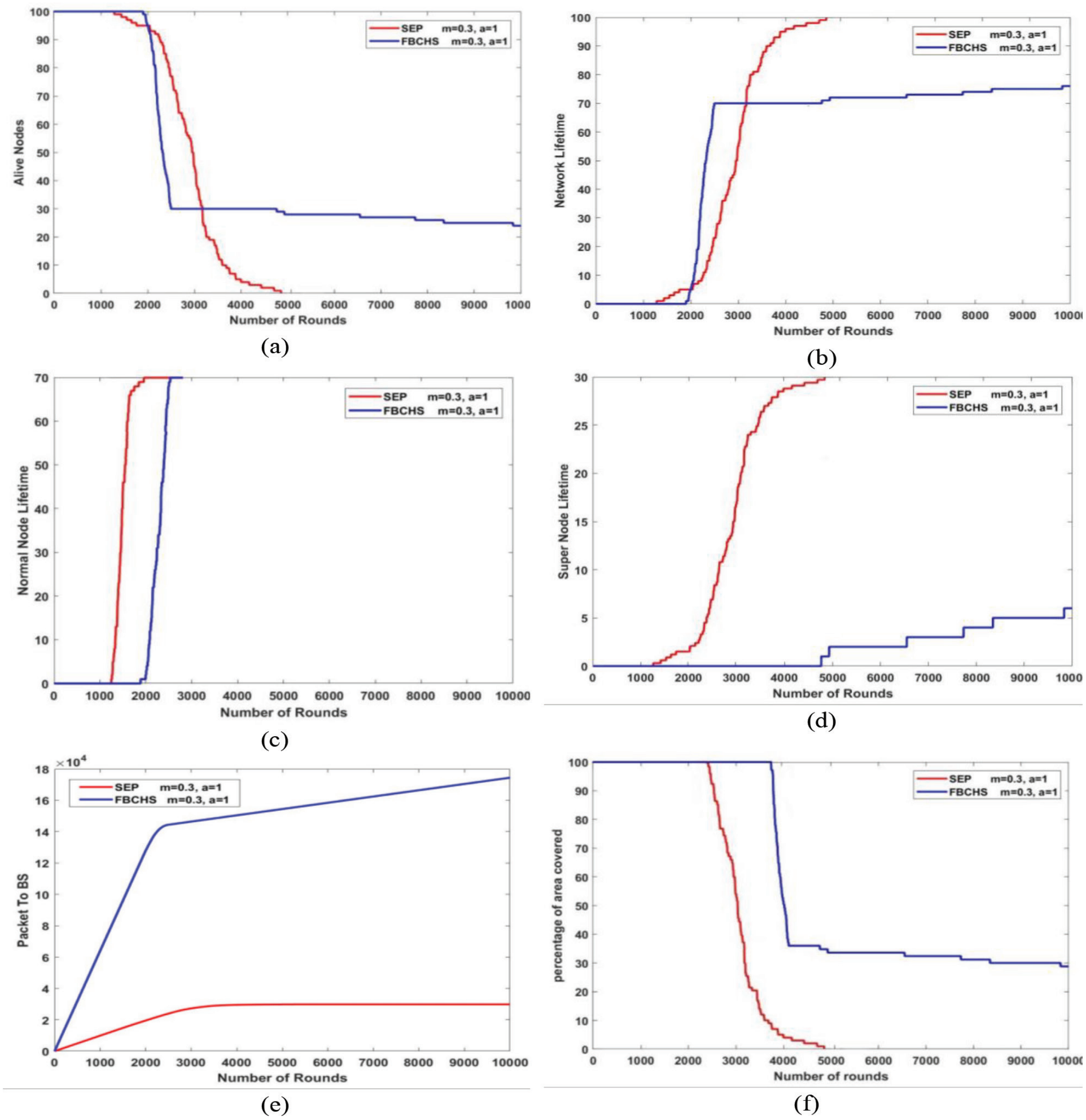


Figure 8. Simulation Performance of SEP and FBCHS. ***(a)*** *The number of Alive Nodes in SEP and FBCHS,* ***(b)*** *Network Life- time of SEP and FBCHS,* ***(c)*** *Lifetime of Normal Node in SEP and FBCHS,* ***(d)*** *Lifetime of Super Node in SEP and FBCHS,* ***(e)*** *Packet Transmitted to BS in SEP and FBCHS,* ***(f)*** *Percentage of the area covered by SEP and FBCHS.*

The RE of Super Nodes in different regions after the 10000 successful rounds of packet transmission is shown in the figure below. Region 1 Super Nodes have 4.4837 joules of remaining energy, region 2 has 4.4861 joules, region 3 has 1.9979 joules, and region 4 has 2.0012 joules of remaining energy. Region 5 Normal Node energy is exhausted completely due to the direct data transmission to BS as shown in Figure 9.

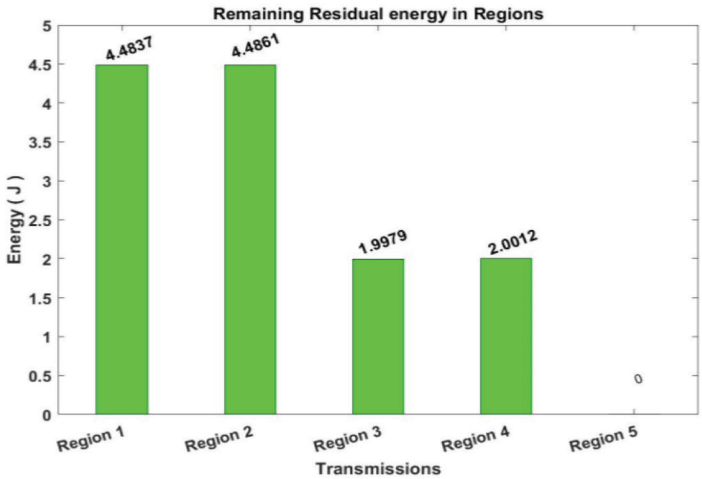


Figure 9. Remaining Energy of Super Nodes in each region

The FBCHS protocol Super Node is alive after 10000 successful transmissions of data. The FBCHS protocol has an overall network lifespan of 13700 rounds, while the SEP protocol network lifetime is 4725 rounds. The red line in Figure 10 indicates that all the nodes in SEP die after 4225 rounds, thus, the lifetime is constant.

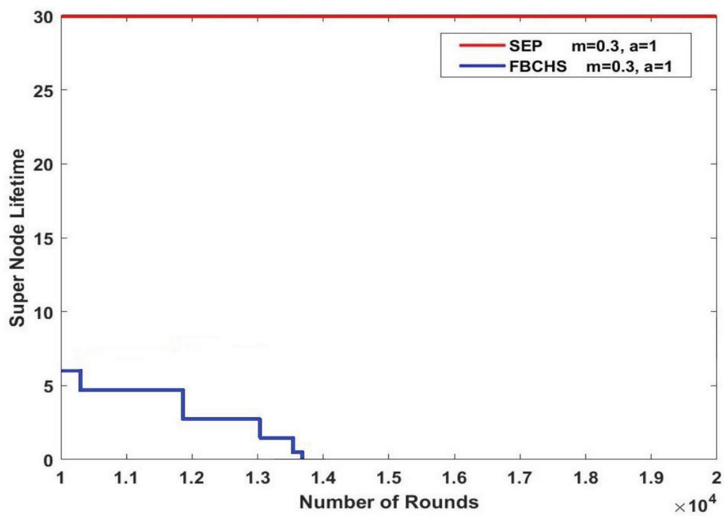


Figure 10. Overall Network Lifetime of SEP and FBCHS

The FBCHS protocol has a much better stability period than the SEP protocol. The SEP protocol trans- mits 3×10^4 packets to BS, afterwards it goes constant because all the nodes' die following a successful round of packet transmission in the network. The FBCHS protocol successfully transmitted 2.1×10^5 pack- ets to BS, which is much more than the SEP protocol as shown in Figure 11.

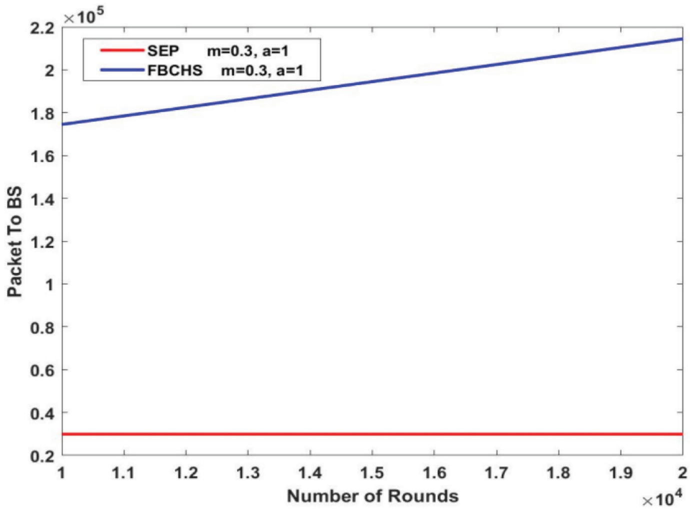


Figure 11. Packet Transmitted to BS in SEP and FBCHS

The FBCHS protocol Super Nodes are alive after 10000 rounds, therefore, they cover the network area much more compared to the SEP protocol (because all nodes die at 4725 rounds in SEP). The FBCHS protocol covers the network area for 13700 rounds as shown in Figure 12.

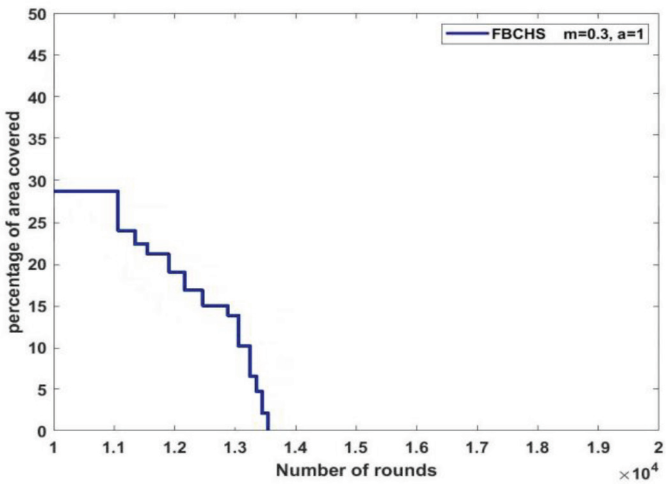


Figure 12. Percentage of the area covered by SEP and FBCHS

6. Conclusion and Future Work

The conservation of node energy is a crucial issue for the sensor network. Cluster-based protocols have proven to be a more effective solution for prolonging the network lifespan. The distribution of Super and Normal Nodes, in the regions, on the basis of energy levels. The proposed FBCHS protocol shows efficient utilization of energy in the network. The SEP protocol assigned weight probability to every node. The FBCHS protocol selected an optimal number of cluster heads per round to enhance the node residual energy of the network. The number of the alive Super Nodes in the FBCHS protocol was significantly higher than in the SEP Protocol. The efficiency of the FBCHS protocol has been contrasted with SEP protocol for stability period, alive nodes number, network lifetime, packet transmitted to BS, percentage of area covered, residual energy of the node, and enhanced network performance. In future research, we will focus on the effectiveness of sensor nodes in a mobility environment and cluster head selection mechanism using a genetic approach, which provides optimal clusters and cluster nodes in the network so that the lifetime of the network will be enhanced.

References

Batra, P. K., and Kushwah, R., 2019. 'Fuzzy Logic based Cluster Head Selection method for Heterogeneous Wire- less Sensor Networks', in 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 86–90.

Behzad, M., 2018. 'M-BEHZAD: minimum distance based energy efficiency using hemisphere zoning with ad- vanced divide-and-rule scheme for wireless sensor networks', arXiv preprint arXiv:1804.00898.

Choudhary, S., et al., 2022. 'Fuzzy Approach-Based Stable Energy-Efficient AODV Routing Protocol in Mobile Ad hoc Networks', in Software Defined Networking for Ad Hoc Networks. Springer, pp. 125–139.

Deepa, O., and Suguna, J., 2020. 'An optimized QoS-based clustering with multipath routing protocol for wireless sensor networks', Journal of King Saud University-Computer and Information Sciences, 32(7), pp. 763–774.

Dwivedi, A. K., and Sharma, A., 2020. 'FEECA: Fuzzy based Energy Efficient Clustering Approach in Wireless Sensor Network', EAI Endorsed Transactions on Scalable Information Systems, 7(27).

Elhoseny, M., et al., 2020. 'Swarm intelligence--based energy efficient clustering with multihop routing protocol for sustainable wireless sensor networks', International Journal of Distributed Sensor Networks, 16(9), p. 1550147720949133.

Faiz, M., and Daniel, A. K., 2020. 'Fuzzy Cloud Ranking Model based on QoS and Trust', in 2020 Fourth Interna- tional Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 1051–1057.

Jayaraman, G., and Dhulipala, V. R., 2021. 'FEECS: Fuzzy-Based Energy-Efficient Cluster Head Selection Algo- rithm for Lifetime Enhancement of Wireless Sensor Networks', Arabian Journal for Science and Engineering, pp. 1– 11.

Kumar, S., et al., 2021. 'Division Algorithm Based Energy-Efficient Routing in Wireless Sensor Networks', Wire- less Personal Communications, pp. 1–20.

Latif, K., et al., 2015. 'Energy hole minimization with field division for energy efficient routing in WSNs', Interna- tional Journal of Distributed Sensor Networks, 11(10), p. 953134.

Latif, K., et al., 2016. 'Energy consumption model for density controlled divide-and-rule scheme for energy efficient routing in wireless sensor networks', International Journal of Ad Hoc and Ubiquitous Computing, 21(2), pp. 130–139.

Maurya, S., and Daniel, A. K., 2015. 'Rbhr: Region-based hybrid routing protocol for wireless sensor networks using ai technique', in Proceedings of Fourth International Conference on Soft Computing for Problem Solving, pp. 37–52.

Mehra, P. S., Doja, M. N., and Alam, B., 2020. 'Fuzzy based enhanced cluster head selection (FBECS) for WSN', Journal of King Saud University-Science, 32(1), pp. 390–401.

Narayan, V., and Daniel, A. K., 2019. 'Novel protocol for detection and optimization of overlapping coverage in wireless sensor networks', Int. J. Eng. Adv. Technol, 8.

Narayan, V., and Daniel, A. K., 2020. 'Multi-Tier Cluster Based Smart Farming Using Wireless Sensor Network', in 2020 5th International Conference on Computing, Communication and Security (ICCCS), pp. 1–5.

Narayan, V., and Daniel, A. K., 2021. 'RBCHS: Region-Based Cluster Head Selection Protocol in Wireless Sensor Network', in Proceedings of Integrated Intelligence Enable Networks and Computing. Springer, pp. 863–869.

Narayan, V., and Daniel, A. K., 2021. 'A novel approach for cluster head selection using trust function in wsn', Scalable Computing, 22(1), pp. 1–13. doi: 10.12694:/scpe.v22i1.1808.

Narayan, V., and Daniel, A. K., 2022. 'CHHP: coverage optimization and hole healing protocol using sleep and wake-up concept for wireless sensor network', International Journal of System Assurance Engineering and Manage- ment, pp. 1–11.

Nazari Talooki, V., Rodriguez, J., and Marques, H., 2014. 'Energy efficient and load balanced routing for wireless multihop network applications', International Journal of Distributed Sensor Networks, 10(3), p. 927659.

Sharma, G., and Kumar, A., 2018. 'Improved DV-Hop localization algorithm using teaching learning based optimi- zation for wireless sensor networks', Telecommunication Systems, 67(2), pp. 163–178.

Shivappa, N., and Manvi, S. S., 2019. 'Fuzzy-based cluster head selection and cluster formation in wireless sensor networks', IET Networks, 8(6), pp. 390–397.

Singh, K., and Daniel, A. K., 2015. 'Load Balancing in Region Based Clustering for Heterogeneous Environment in WSNs Using AI Techniques', in 2015 Fifth International Conference on Advanced Computing & Communication Technologies, pp. 641–646.

Sirsikar, S., and Chandak, M., 2018. 'Efficient Clustering using Concentric Rings and Rectangular Region For- mations in Wireless Sensor Networks', International Journal of Applied Engineering Research, 13(6), pp. 3483–3491.

Tang, W., Zhang, K., and Jiang, D., 2018. 'Physarum-inspired routing protocol for energy harvesting wireless sensor networks', Telecommunication Systems, 67(4), pp. 745–762.

Wan, C., and Du, S., 2011. 'Improvement and simulation of leach in wireless sensor networks', Jisuanji Yingyong yu Ruanjian, 28(4), pp. 113–116.

Yang, L., et al., 2018. 'An unequal cluster-based routing scheme for multi-level heterogeneous wireless sensor net- works', Telecommunication Systems, 68(1), pp. 11–26.

Authors's Biography

	Vipul Narayan received the B. Tech degree in Computer Science & Engineering from M.D.U. Rohtak, India in 2012 and M. Tech degree in Computer Science Engineering from M.M.M. University of Technology, Gorakhpur, India in 2016. Now he is research scholar in department of Computer Science & Engineering from 2018 to till date. His current research interests in Wireless Sensor Network. He has published 16 papers in International Journal and International conferences.
	A.K. Daniel is currently working as a Professor in the Department of Computer Science and Engineering, M.M.M. University of Technology, Gorakhpur and has an experience of over 35 years in teaching and research. He is a senior member of IEEE, ACM, CSI and various others. He had published over 100 papers in the national as well as international conferences and journals. He has guided a number of research scholars of post graduate level. His research interest includes wireless communication, artificial intelligence and mobile ad hoc network protocols.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 309-329
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.27902

An Optimized Deep ConvNet Sentiment Classification Model with Word Embedding and BiLSTM Technique

Roop Ranjan[a] and AK Daniel[b]

[a, b] Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh, India, 273001

roop.ranjan@gmail.com, danielak@rediffmail.com

KEYWORDS

sentiment classification; deep learning; long short term memory networks; emotion analysis; FastText

ABSTRACT

Sentiment Classification is a key area of research in natural language processing and it is frequently applied across a range of industries. The goal of Sentiment Analysis is to figure out if a product or service received a negative or positive response. Sentiment Analysis is widely utilized in several commercial fields to enhance the quality of services (QoS) for goods or services by gaining a better knowledge of consumer feedback. Deep learning provides cutting-edge achievements in a variety of complex fields. The goal of the study is to propose an improved approach to evaluating and categorising sentiments into different groups. This study proposes a novel hybridised model that combines the benefits of deep learning technologies Dual LSTM (Long Short-Term Memory) and CNN (Convolution Neural Network) with the word embedding technique. In addition, attention-based BiLSTM is used in a multi-convolutional approach. Standard measures were used to verify the validity of the proposed model's performance. The results show that the proposed model has an enhanced accuracy of 97.01%, which is significantly better than existing models.

1. Introduction

Sentiment Analysis is a set of linguistic operations which apply to digitized texts, such as publications and comments from social networks, as well as press articles, as part of the autonomous processing of natural language. Its goal is to determine the sentiment represented in a text and predict its polarity (positive or negative) toward a specific subject. Sentiment Analysis is a technique for the automatic extraction of sentiment data from unstructured texts. Several fields, including machine learning, natural language processing (NLP), and data mining, use Sentiment Analysis (Liu et al., 2012). Researchers are becoming increasingly interested in the different methods of analysing this massive amount of data for extraction of the views stated on a specific topic. Sentiment Analysis is used by corporations to learn how their customers feel about their products and services. People may gain feedback on products, services, events, and other topics thanks to the widespread use of social media platforms (Pang et al., 2008). People can express their opinions and share their reviews on social media such as Twitter, Instagram, blogs, and news sites. The capability to notice and respond to negative customer feedback is critical to a company's success. (Kim et al., 2017).

Statistical machine learning approaches are great for simple Sentiment Analysis, but they are not suitable for more complex text classification problems. (Neethu et al., 2013).

Word embedding is a technique for representing written words as a set of numeric values or vectors. It generates similar vector representations for words with similar meanings. The distributed representation-based word embedding systems Word2Vec (Lebret et al., 2013) and Glove (Pennington et al., 2014) are now the most widely utilised. Deep learning approaches are currently preferred over machine learning methods such as Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Random Forests, which are often used for sentiment categorization, according to a recent study (Rehman et al., 2019).

RNN is a deep learning technique that is commonly used in natural language processing (NLP) for predicting the upcoming word on the basis of the previous words in the given sentence. Convolutional Neural Networks (CNN) are a type of artificial neural network that can accurately recognise information in a variety of situations (Simonyam et al., 2014). Due to the cost-effectiveness of large manufacturing of efficient Graphics Processing Unit cards and the large size of datasets, deep learning has emerged as a superior method to machine learning in sentiment classification applications (Chollet et al., 2017). Compatibility between text representation techniques and their algorithms is critical in many text-classification tasks. Therefore, it becomes feasible for providing high-performance classification with the assistance of a suitable representation of texts and a classification model.

With the advancements in cognitive computing and artificial intelligence, deep learning techniques have emerged as an increasingly popular method for Sentiment Analysis in several fields (Chen et al., 2018; Hwang et al., 2017). As a result, a growing number of researchers use neural networks to automatically build feature representations from text sequences. Currently, the prominent deep learning algorithms for sentiment categorization include RNN, CNN, and LSTM. Huang et al., 2015, proposed combining bidirectional LSTM (BiLSTM) and CRF models to get superior outcomes by using both backward input and forward input characteristics. Ma and Hovy, 2016, proposed a combined approach using CNNs and LSTM-CRF to identify entities.

Each deep learning approach has a distinctive property that is actively used to achieve a particular objective or assess datasets. The text representation technique, which translates texts into numeric form, has both benefits and disadvantages. To resolve the difficulties and obtain higher accuracy, a combined approach is critical. It would require ideal text depiction and an appropriate machine learning algorithm for emotion analysis. In the numeric format representation of user sentiments, any text representation approach (e.g., word embedding, character embedding) is not fully complete. To address these issues and to evaluate a dataset of tweets from various Indian airline services, a new hybrid model is proposed using deep learning techniques that employ multiple representations of data.

To achieve the objective, the main contribution of this research is to use the advantages of different deep learning techniques i.e., CNN and DualLSTM followed by another layer of CNN. A CNN layer is used for feature extraction, which is then passed to a bidirectional LSTM to obtain long-duration dependencies. Following these steps, the second layer of CNN is used to begin learning the retrieved properties from scratch. This process simplifies the learning of the features, and the machine gains a better knowledge of the classes. Word embedding was used in our research to represent words. Furthermore, word vectorization is combined with the orientation score characteristics as the sentiment characteristics set. The embedding parameters were merged and put into a CNN layer and a DualLSTM network layer. To test the validity and effectiveness, the model is implemented on five different sets on the basis of the variation in the length of the reviewed text. The proposed model has significantly improved results in comparison to previous models.

The major contribution of the research are as follows:

1. Word embedding techniques Word2Vec, Glove, and FasText have been utilised to render tweets as numerical vectors. The models are pre-trained word vectors which can capture word semantics and are trained on a huge collection of words. The research that has been carried out has focused on several models of vectors of words to test the performance of the model.
2. By combining the framework of CNN with DualLSTM, a multi-stage ConvNet (CNN) model-based technique for text sentiment classification is presented. The ConvNet model obtains localized features from word embedding, the Bi-LSTM model captures long-distance dependencies, and the outcome of the classification are the classified features.
3. The efficiency of the proposed model is compared with several deep learning methods and the outcome proves the effectiveness of the model.

The rest of the research paper is structured as follows: Section 2 provides an overview of state-of-the-art research, particularly relating to the application of Sentiment Analysis in relevant fields. The techniques that have been involved in this proposal are discussed in Section 3. Section 4 provides detailed information on the dataset and classifier models that were utilized in the experiments. The experimental findings are outlined in Section 5, which is followed by a discussion in Section 6. In the last section, the article is concluded and future lines of research are described.

2. Related Work

Sentiment Analysis has received increasing attention from researchers as a result of the growth of NLP, and various implementations of word-level embedding have been carried out. State-of-the-art research has discovered that hybrid algorithms produced improved classification results. Jiang *et al.*, 2016, created a text representation strategy using bag-of-words based on sentiment topical terms, which included a deep neural network, context information, and sentiment topic terms that performed significantly better in Sentiment Analysis. Whereas Rezaeinia et al., 2019, presented an enhanced word embeddings technique based on POS tagging methodology and lexicons of sentiment that outperformed pre-trained embeddings of words utilized in Sentiment Analysis. To improve the performance of sentiment classification, Liu et al., 2016, presented a model in which machine learning was coupled with deep learning. The efficiency of the presented method was shown using datasets with Turkish and Chinese languages in their research. Pham et al., 2018, presented an approach that combined several CNNs and was centered on embeddings of words from Glove, Word2Vec, and it performed well in aspect-level sentiment classification tasks. Han et al., 2019, developed a hybrid neural network approach for document representation that included user and product information and used CNNs and LSTM.

An innovative deep learning methodology with hybrid CNNs and Bi-LSTM features was demonstrated in Wint et al., 2018, that coupled the power of CNNs with Bi-LSTM. The authors achieved unique vectors of features provided as input to the LSTM layer using both separate pre-trained vectors of words. The BERT model for representing text, proposed by Devlin et al., 2019, may better depict the modifying associations in texts and performed well in Sentiment Analysis tests. By combining a neural topical approach into representations of word-level semantics, Liu et al., 2019 proposed latent topic information of the given text, as well as a novel topic-based attention method for texts to look at the syntactic of words using the topics given for word association. BiLSTMs were used in another research Guo et al., 2018, to extract dependency knowledge from vocabulary and the position of the sentence. In the proposed hybrid strategy, BiLSTM and CNN were used to produce n-gram features from text categorization by applying multiple CNNs to given LSTM outputs.

In another approach, Zhou and Long, 2018, performed the classification of texts from reviews of Chinese products using the Bi-LSTM approach, with the help of CNN for extracting features using word embedding. The combination of CNN and BiLSTM for classification in the experiments led to superior classification results in comparison to using CNN and BiLSTM separately. In another research (Sun et al., 2014), a multilayer CNN with LSTM was utilized for analyzing the sentiments of users on a given dataset with the help of the application of social media from Tibetan. The characteristics were retrieved using the assistance of a three-layered CNN. The collected features were sent to a two-layer LSTM network as input. This hybrid model, based on deep learning, outperformed CNN and LSTM according to the results. A hybrid Bi-LSTM model based on attention-mechanism is proposed in another research (Zheng et al., 2019). The proposed model efficiently integrated BiLSTM and CNN for the classification of text using Word2Vec with an attention mechanism.

In addition, Fan et al., 2017, utilized a single model of CNN and a dual-stacked LSTM to analyse Indian tweets on a sequential basis. The characteristics produced from the CNN layer are fed into the LSTM network in their hybrid experiments. Like earlier hybrid research, the researchers utilized CNN to extract characteristics. Zhou et al., 2016, proposed a dual language representational learning model using attention, in which the scattered semantics of several text documents in both target and source languages were learned. Documents are modelled with LSTM networks. They also proposed a hierarchy-based attention methodology for multilingual LSTM networks, which outperformed the benchmark data set.

Furthermore, the Convolutional Neural Network (Xu et al., 2019) method might efficiently extract local information. Also, the work has been done utilizing CNN to handle the issue of orientation analysis. AF-LSTM was proposed by Tay et al., 2018. To represent context and aspect terms at the word level, Atrial Fibrillation-LSTM (AF-LSTM) employs circular correlation and convolution. Through the information fusion procedure, the targeted knowledge is then integrated into the representation of sentences. Yu et al., 2019, utilized a multi-way Gated Recurrent Unit (GRU) combined with an attention-based mechanism to classify brief text in e-commerce reviews, with promising results. These approaches demonstrate that when deep learning is coupled with an attention mechanism short text Sentiment Analysis can improve.

In another research, Huang et al., 2018, presented an AOA model i.e., Attention-over-Attention (AOA) model. With the help of two fine-grained attention mechanisms, the AOA model extracts text-level association between aspects and contexts that allow the sentence representation to focus automatically on the elements that belong to sentences that are significantly more important for the aspect terms expressions. The TDLSTM model is proposed by Tang et al., 2016. In the proposed model the sentence was split into two sections with aspect. The hidden state of both sections was modeled using dual long short-term networks (LSTM). After that, the two portions were merged to create a specified target representation that was then put into the activation softmax algorithm for emotion categorization. Baziotis et al., 2017, utilized LSTM coupled with an attention-based model to assign weights to decisive words via the attention-based mechanism, which improved the effect of keywords of emotion sentences and obtained significantly better results in the categorization of emotions in Twitter reviews.

Although several researchers have proposed deep neural network architectures for sentiment classification problems based on the CNN or RNN, few researchers have comprehensively examined the performance of several classification models based on deep learning methods. Seo et al., 2020, proposed an analytical comparison of several sentiment classifications based on deep learning model architectures to extract useful significance for the development of the sentiment classification approach. For the identical model structure, input of word-level produced better results for classification than the input based on character-level. In another study, Hu et al., 2018, showed that models built using deep learning outperformed standard techniques, such as algorithms based on dictionary methods, the Naive Bayes, or SVM. They did not, however, disclose quantifiable performance metrics such as F-Measures or level of accuracy. Dzikienė et al., 2019, presented a performance comparison of conventional machine learning methodologies Naïve-Bayes Multinomial, Support Vector Machine and deep learning (LSTM and CNN) approaches on the Lithuanian internet comments dataset. Features based on morphological, character information, conventional machine learning approaches were applied. The deep learning approaches were applied on both Word2Vec and FastText embeddings. In their experiments, LSTM outperformed SVM and Naïve Byes multinomial methods. Yin et al., 2017, compared the results of LSTM, CNN, and GRU sentiment classification. However, the utility of their experimental outcomes was of less use since they could not focus on enough structure of differences in the model, and their conclusive result was based on the experimental results of only one dataset.

RNN was utilized by Socher et al., 2013, to tackle text categorization difficulties. A model called Sentiment Treebank was introduced that surpassed all earlier techniques in terms of various criteria when the training was performed on the fresh tree-bank, which could represent the impacts of negativity accurately. Yang et al., 2016, established a hierarchy-based text categorization strategy based on the attention-based mechanism that effectively captured the text's main sentiment information. Huang et al., 2015, suggested combining bidirectional LSTM (BiLSTM) and CRF models to produce superior outcomes by using both backward and forward input characteristics. Sentic-LSTM was developed by author to explicitly incorporate explicit and implicit information, and an extended version of Sentic-LSTM was presented to deal with a combined work involving aspect detection of target-dependent and aspect-based classification of polarity (Ma et al., 2018). The authors presented a refinement method of word vector that improved all word vectors. By improving pre-trained vectors of the word and using intensity ratings of sentiments provided by sentimental lexicons, the proposed model performed better in Sentiment Analysis (Gu et al., 2018). By training a huge corpus of text, Peters et al., 2018, proposed a textual representation strategy using a deep learning model, the researchers built a text depiction framework in the English language that incorporated grammar and sentiment elements.

Looking at the studies described in these papers, it is evident that a variety of methods have been implemented for the classification of sentiments using deep learning techniques. The approaches and research described above were primarily employed for extracting semantic information for features from the sentence dimension while ignoring the information-based features of the dimension of the word vector. In this research, two Convolutional layers (ConvNet) are used with an attention-based BiLSTM layer to extract the semantic information of the local characteristics of the word vector in the word insertion dimension. This research also uses max pooling to achieve significantly comprehensive local feature information. Section 3 discusses the proposed model in detail.

3. Proposed Method

The core idea behind the proposed mode is to use two distinct deep neural networks, namely ConvNet and DualLSTM. Then a hybrid ConvNet + DualLSTM + ConvNet model is used to detect users' view orientation towards the services represented in Comments in tweets as shown in Table 5. To reduce unnecessary information in sentences, it is required to create a base sentiment dictionary with negative, positive, and neutral texts that contain only relevant sentiment words that are included manually, as well as a set of rules based on a grammatical sentence that includes associated words and degree adverbs. Among Word2Vec, Keras and FastText; FastText embedding has shown the most optimized performance. Therefore FastText is chosen as the word embeddings method in the proposed research. Further, the hybrid model is proposed because ConvNet extracts local features from comments, DualLSTM captures contextual information from both directions as well as long-range dependencies, and the hybrid model combines the benefits of both complimentary ConvNet and DualLSTM architectures.

In this section, a detailed explanation of the proposed model is discussed. The proposed model consists of the following layers:

3.1 Pre-Processing Layer
3.2 Word Embedding Layer
3.3 ConvNet Layer(Multi)
3.4 Pooling Layer
3.5 DualLSTM Layer
3.6 Attention and Dense Layer
3.7 Output Layer

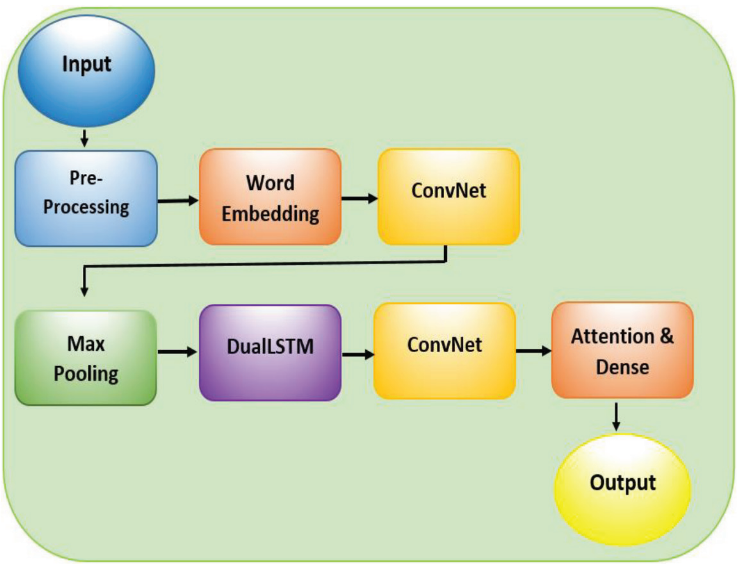


Figure 1. Proposed Model of Sentiment Classification

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 309-329
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

3.1. Pre-Processing Layer

Raw information usually contains words or symbols that computers cannot comprehend, therefore it is required to clean the data and restructure it in an understandable format, and therefore data cleaning is the important stage in NLP. The pre-processing of input texts in Figure 1 was implemented on the dataset to obtain excellent classification performance in categorizing text review data. Before the word-based representation phase (Word Embedding), these preprocessing steps are used to reduce unwanted contents and convert the dataset into a usable form. Initially, all texts in the dataset are converted to lower cases. Further, irregular spacing between words is trimmed to a single space. Punctuations, numerals, and unformatted characters in tweets were eliminated. Grammatical Error Correction (GEC) by NLP-Progress is used to correct all grammatical mistakes like spellings, grammar, punctuations.

3.2. Word Embedding Layer

One of the most important stages in the text classification process is the accurate representation of texts. In this study Word2Vec, FastText Embedding and Glove Embedding are used.

The steps for converting input text to Word Vector Representation are as follows:

- The method takes as input a text labelled as C with m tokens (words) with each token transferred to its corresponding word vector representation *s*. The text is reorganized in a word embedding concatenated sequence.
- $C = [s_1; s_2, s_3;::: :s_m]$, where s_i is the embedding of the word for the i^{th} word, which is projected to a given vector $s_i \in S^d$.
- A sentence $\in S^{dxm}$) matrix is created for each text input, where *d* indicates the embedding dimension and has been defined as the length of the sentence.
- The sentence's matrix is now transferred to the ConvNet layers, where it will be processed further.

3.3. ConvNet Layer

As illustrated in Figure 2, the architecture of a Convent comprises an input layer, an output layer, and five different hidden layers. The input layer accepts a textual message that has been padded to a predetermined length of words, followed by a word embedding layer. The attention layer follows the word embedding layer, to extract high-level feature vectors. The attention layer is a sub-unit made up of context vectors that line up the source input with the goal output. Figure 6a shows an illustration of the attention mechanism in the upper right corner. The SpatialDropout1D (for dropout) layer uses feature vectors as inputs derived from the attention layer. On top of the dropout layer, a ConvNet layer with convolution filters and a ReLU activation function is applied. The details of ConvNet model are shown in Figure 2.

- The ConvNet Module uses a convolutional based operation «*», between the matrix of text $C \in S^{dxm}$
- A filtering matrix $P \in S^{mxk}$, which provides an output as matrix *OP*, is termed as a features map.
- The features map is learned as per the following equation:

$$O_{ab} = (C * P) = k(D \circ t_{a:a+n-1,b+d-1} + y) \quad (1)$$

Where y represents the biasing vector, D represents the weighted matrix, and k is the convolutional operation's nonlinearity activation function. A nonlinearity function Rectified Linear Unit (ReLU) is applied here for speeding up the process of training and validation and provides better results.

3.4. Pooling Layer

The pooling layer receives the output of the convolutional layer. The convolutional layer minimizes the content representation even further by selecting the maximum value obtained from a pool of values and eliminating the irrelevant data. The procedure of pooling is represented as follows:

$$V_{a,b} = \max\left(O_{a+b-1,b+d-1}\right) \tag{2}$$

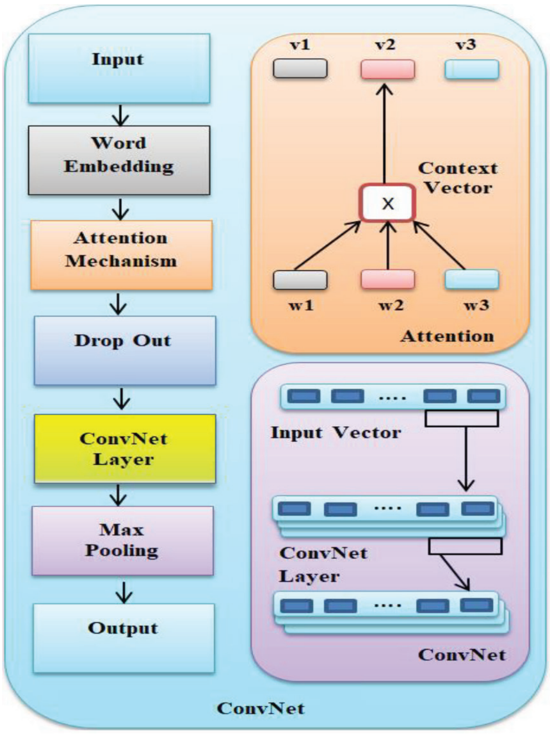


Figure 2. Architecture of ConvNet

3.5. DualLSTM Layer

To make precise predictions, the model must understand the long-duration dependence on text data. Since the convolutional layer lacks this capacity, DualLSTM has been used to incorporate this component into the proposed model. The model learns from the data in both directions i.e., left-to-right and right-to-left using DualLSTM. As a result, the DualLSTM layer enhances classification accuracy. In the Bidirectional LSTM, there are two autonomous LSTMs i.e., ahead LSTM and backward LSTM.

The hidden state «h» is computed by forwarding LSTM using the hidden previous state «h_{t-1}» and the input vector «z_t» whereas the backward LSTM uses the hidden future state «h_{t+1}» and the vector input «z_t» to compute the hidden state «g».

Finally, using the following Equation 3, both direction vectors(backward and ahead) have been merged as the last state(hidden) in the DualLSTM to create a series of output for vectors(hidden) $H= [h_1; h_2; h_3;......h_n]$.

$$\overleftrightarrow{h} = \overrightarrow{h} \oplus \overleftarrow{h} \tag{3}$$

The following equations are used to implement the DualLSTM cell.

Forward LSTM:

$$R = \begin{bmatrix} h_{t-1} \\ z_t \end{bmatrix}$$

$$\sigma f p_t = (W_{fp} \cdot R + y_{fp}) \tag{4}$$

$$\sigma i p_t = (W_{ip} \cdot R + y_{ip}) \tag{5}$$

$$\sigma o p_t = (W_{op} \cdot R + y_{op}) \tag{6}$$

$$c_{pt} = m \odot c_{pt\text{-}1} + ip_t \odot \tanh\,(W_{cp} \cdot R + y_{cp}) \tag{7}$$

$$h_t = op_t \odot \lambda(c_{pt}) \tag{8}$$

Backward LSTM:

$$X = \begin{bmatrix} h_{t+1} \\ z_t \end{bmatrix}$$

$$\sigma f p_t = (W_{fp} \cdot R + y_{fp}) \tag{9}$$

$$\sigma i p_t = (W_{ip} \cdot R + y_{ip}) \tag{10}$$

$$\sigma o p_t = (W_{op} \cdot R + y_{op}) \tag{11}$$

$$cp_t = k_{pt} \odot c_{pt\text{-}1} + ip_t \odot \tanh\,(W_{cp} \cdot R + y_{cp}) \tag{12}$$

$$h_t = op_t \odot (cp_t) \tag{13}$$

Where weight matrices are represented by W_{fp}, W_{ip}, W_{op}, and y_{fp}, y_{ip}, y_{op} as associated biases, which are the input gate parameters, forget gate parameters, and output gate parameters respectively.

sigmoid function for activation is represented by σ.
symbol-wise multiplication is denoted by $\odot$.
g_t is representing the vector in hidden state and the input vector is denoted by z_t.
λ is a tangent function.
The current state is denoted by cp_t, the previous state is shown by cp_{t-1} and future state is represented by cp_{t+1}

3.6. Attention & Dense Layer

There are certain words in a statement that are irrelevant for polarity detection but on the other hand, some words are decisive. The attention-based mechanism is used to draw attention to informative content. Therefore, this layer was created to automatically extract the significant terms.

Eq. 14 is used to calculate the word significance vector e_t. It uses the whole DualLSTM hidden states *h,* as input to the attention layer. *W* stands for weight, *y* for bias, and *tanh* for activation function.

$$e_t = \tanh\,(W_h h_t + y_n) \tag{14}$$

Finally, to create the output of the attention mechanism, Eq.15 is used to calculate a weighted summation.

$$cp_t = \sum_{t=1}^{m} n_t h_t \tag{15}$$

The attention layer output $cp = [cp_1, cp_2, cp_3, \ldots\ldots cp_m]$ is provided as input for the following layer.

Further conversion of the matrix of context retrieved from the preceding layer to a vector context that provides the input for the classification layer's final stage is performed. The following Eq. 16 is used to execute the flatten layer operation.

$$f = [cp_1 * cp_2 * cp_3 \ldots\ldots * cp_T] \tag{16}$$

3.7. Output Layer

The proposed approach for determining the class of sentiments in terms of negativity, positivity, or neutrality has reached its conclusion. The output of the flatten layer is provided to a softmax activation function, that calculates the likelihood of the sentiment classification. The final output is computed as:

$$O_j = \sum_{i=1}^{m} w_i * k_i + y \tag{17}$$

4. Data Acquisition and Experimental Setup

4.1. Data Acquisition

The proposed model has been tested on a dataset acquired from user tweets on Twitter regarding Indian Airlines between June 1st and August 31st, 2021. The dataset contains 24,235 tweets. Domestic flights began after the second wave of Covid-19, hence the data was obtained during that time period to test the response of the travellers to their services. Therefore, this research provides an insight into the service quality of Indian Airlines after Covid-19. A Rest API-based tool named Tweepy has been used. Three sentiment classes have been represented i.e., positive, negative, and neutral. The dataset has been divided into two distinct subsets i.e., Training Set and Validation Set. 75% of the tweets (18176 tweets) are grouped as the Training Set and 25% (6059) of the tweets are grouped into the Validation Set.

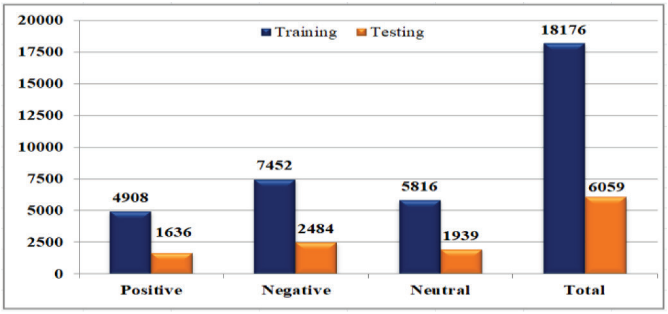


Figure 3. Classified number of Tweets in the dataset

Figure 3 shows the graphical representation of review classification in different categories

Table 1. Categorization Details of the Tweet Dataset

Dataset Size	Positive	Negative	Neutral
24235	6544	9936	7755

Table 2. Categorization of the Training Set

Dataset Size	Positive	Negative	Neutral
18176	4908	7452	5816

Table 3. Categorization of the Validation Set

Dataset Size	Positive	Negative	Neutral
6059	1636	2484	1939

The total number of tweets in the dataset, tweets in the training set, and tweets in the validation set are provided in Tables 1, Table 2, and Table 3, respectively.

4.2. Experimental Setup

Experiments have been conducted utilising Google's services. Google offers a cloud-based file storage service called Google-Drive, which we used to store our data. The Google Colab system was used for this research, which is a free cloud-based tool provided by Google for Machine Learning developers that is built on Jupyter notebook for undertaking machine learning research using Python. Experiments with the proposed model have been carried out using the Keras API with Tenserflow as a backend.

4.3. Parameter Settings

Hyper-parameter optimization must be used to get high model performance. Underfitting and overfitting are also avoided using hyper-parameter settings. To improve accuracy, the randomised search strategy was applied. The hyper parameters were identified as Size of Kernel to be 5, Embedding Dimensions were 300(Keras), Size of Filter was128. The activation function used here was ReLu. Batch size for each epoch was 128 with a total number of 50 epochs. Learning rate was set at 0.01 with Adam Optimizer used for optimization.

4.4. Performance Metrics

The performance of the proposed system was assessed using the standard evaluation matrix illustrated in Figure 4.

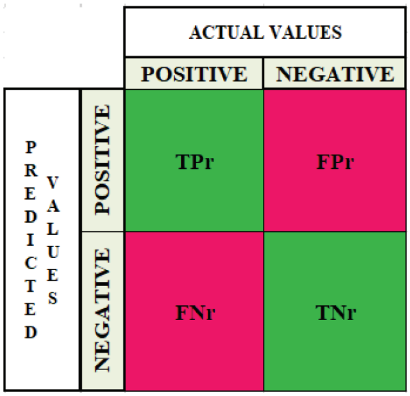


Figure 4. Standard Evaluation Parameters

The standard validation parameters are described below:

- True Negative (TN) - These are accurately forecasted negative outcomes, demonstrating that the value of actual class is zero and the outcome of the anticipated class is zero i.e. correct prediction of negative classes.
- True Positive (TP) - TP are observed positives that are accurately predicted and indicate that the outcome of the actual class and the outcomes of the expected class are positive i.e. correct prediction of positive classes.

False negative and false positive occur if actual class is different from the anticipated class.

- False Positive (FP) – Positive observations of the predicted class and negative observations of the actual class, indicating inaccurate prediction of positive classes.
- False Negative (FN) - When the actual class is positive while the anticipated class is negative i.e. errorneous prediction of negative classes.

Using these standard parameters the following rules were implemented for the evaluation of the effectiveness of the proposed hybrid model:

Precision (P_r) = The proportion of correctly anticipated positive outcomes to total projected positive outcomes is referred to as precision.

$$P_r = \frac{TP_r}{TP_r + FP_r} \quad (18)$$

Recall (R_r) = the proportion of correctly predicted positive outcomes in the positive class to the total number of observations.

$$R_r = \frac{TP_r}{TP_r + FN_r} \quad (19)$$

F-Measure (F_r) = F-Measure is defined as the average of Recall and Precision. Both false negatives and false positives are taken into account when calculating the final score.

$$F_r = \frac{2*(P_r * R_r)}{P_r + R_r} \quad (20)$$

Accuracy (A_r) = The most essential performance metric is accuracy, which is simply the fraction of predicted observations that match all observed.

$$A_r = \frac{TN_r + TP_r}{TN_r + TP_r + FN_r + FP_r} \quad (21)$$

5. Experimental Results and Analysis

This section provides detailed information on the results and a comparative study of performance.

5.1. Comparison of Word Embeddings

The proposed model has been used for the classification of sentiments. The overall performance of the word embedding is assessed using a weighted average of P_r, R_r, and F_r.

The experiments have been performed to evaluate the overall efficiency of classification on word-2vec, Glove and the FastText embedding methods on the used dataset. The performance of efficiency of different word embeddings is represented in Table 4. Glove embedding was observed to be less efficient and lower accuracy as compared to the other two methods. The FastText method of embedding attained optimized performance.

Table 4. Performance of Word Embeddings

Word Embedding	Weighted Average(%)			Acc(%) *(Ar)*
	Pr	***Rr***	***Fr***	
Word2Vec	9236	92.48	92.42	92.83
Glove	9332	94.15	93.73	94.22
FastText	95.92	95.56	95.74	96.32

Compared with Word2Vec, Glove embedding has shown improved efficiency by 0.96%, 1.67%, 1.31% and 2.49% on P_r, R_r, F_r and A_r Furthermore, the efficiency of Word2Vec is comparable to hat of FastText Embedding and the improvement has been observed to be 2.56%, 3.08%, 3.32%, 3.49% in terms of precision, recall, F-measure, and accuracy parameters respectively. When comparing Keras and FastText, it has been noticed that the FastText method has performed better than Word2Vec Embedding on the given dataset in the text, if used as mathematical symbols.

5.2. Performance Comparison with Deep Learning Methods

CNN vs Proposed Model: In this experiment, the classification of emotion exhibited in textual airlines reviews using the proposed model is compared to that of a single layer CNN model (Figure 5). In comparison to the suggested Attention-based proposed model, a single layered CNN model produced unsatisfactory outcomes (*Pr*: 86.35%, *Rr*: 86.98, and *Fr*: 86.66% for positive classification, *Pr*: 86.01%, *Rr*: 86.88 and *Fr*: 86.44% for neutral classification, *Pr*: 85.32%, *Rr*: 84.33 and *Fr*: 84.82% for negative classification and *Ar*: 86.35% accuracy). The reason for CNN's poor performance is that it is unable to retain the text's sequencing order required for the text categorization problem for keeping record of the details of ordering to provide improved classification results (*Online*).

Table 5. Performance Comparison of Deep Learning Methods

Models with FastText	Class 1			Class 2			Class 3			Acc% *(Ar)*
	Pr	***Rr***	***Fr***	***Pr***	***Rr***	***Fr***	***Pr***	***Rr***	***Fr***	
CNN	86.35	86.98	86.66	86.01	86.88	86.44	84.32	84.22	84.27	86.35
BiLSTM	88.15	88.98	88.56	90.11	89.32	89.71	85.32	84.33	84.82	89.32
CNN-BiLSTM	89.21	88.32	88.76	83.35	83.98	83.66	87.65	87.33	87.49	90.35

Table 5. Performance Comparison of Deep Learning Methods ((continued))

Models with FastText	**Class 1**			**Class 2**			**Class 3**			**Acc% (Ar)**
	Pr	***Rr***	***Fr***	***Pr***	***Rr***	***Fr***	***Pr***	***Rr***	***Fr***	
BiLSTM-Attention	87.65	86.32	86.98	93.26	93.14	93.20	90.25	91.25	90.75	92.63
Proposed Model	94.65	94.12	94.38	96.98	96.25	96.61	95.12	95.56	95.34	97.01

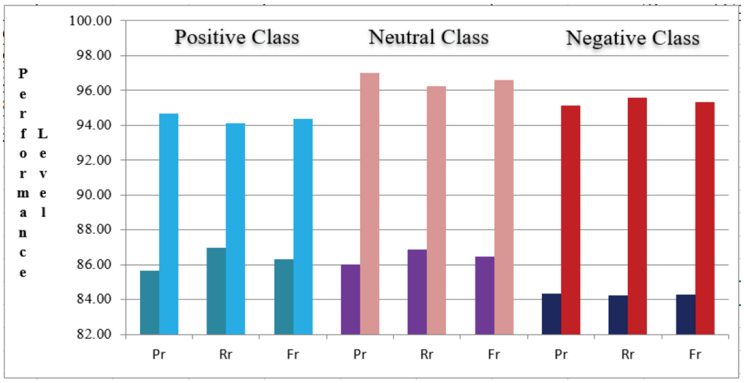


Figure 5. Performance Comparison of CNN and proposed model

BiLSTM vs. Proposed Model: The performance of the Bi-LSTM model was analyzed and comparison was performed with the proposed model in the next experiment (Figure 6). The Bi-LSTM delivers lower outcomes (*Pr*: 88.15%, *Rr*: 88.98%, Fr: 87.95% for positive classification, *Pr*: 85.32%, *Rr*: 84.33, *Fr*: 84.82% for neutral classification, *Pr*: 90.11%, *Rr*: 89.32, *Fr*: 89.71% for negative classification, *Ar*:89.32% accuracy) than the suggested model. The reason for degradation in the performance of the Bi-LSTM model is that the process is costly in terms of memory utilization due to double LSTM cell requirements (Zhu et al., 2018).

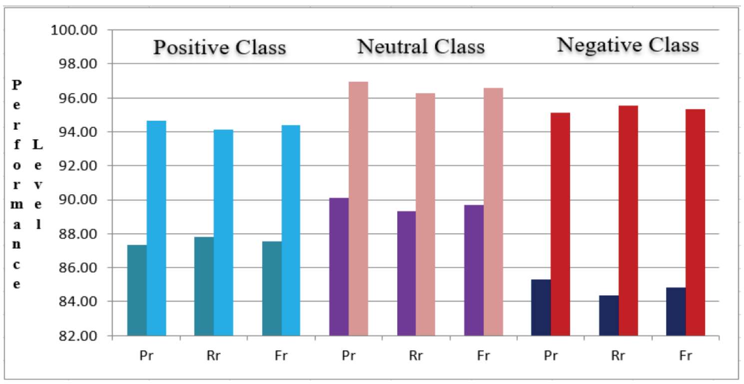


Figure 6. Performance Comparison of BiLSTM and Proposed Model

CNN-BiLSTM vs. Proposed Model: Furthermore, the research was performed to compare the performance results obtained from the proposed model to the CNN-BiLSTM model (Figure 7). In comparison to the suggested technique, the CNN-BiLSTM delivers poor outcomes (Pr: 89.21%, Rr: 88.32%, Fr: 88.76% for positive classification, *Pr*: 83.35%, *Rr*: 83.98, *Fr*: 83.66% neutral classification, *Pr*: 87.65%, *Rr*: 87.33, *Fr*: 87.49% for negative classification and *Ar*: 90.35% accuracy). When compared to the suggested model, BiLSTM performs poorly because it lacks the attention mechanism (Yao et al., 2017).

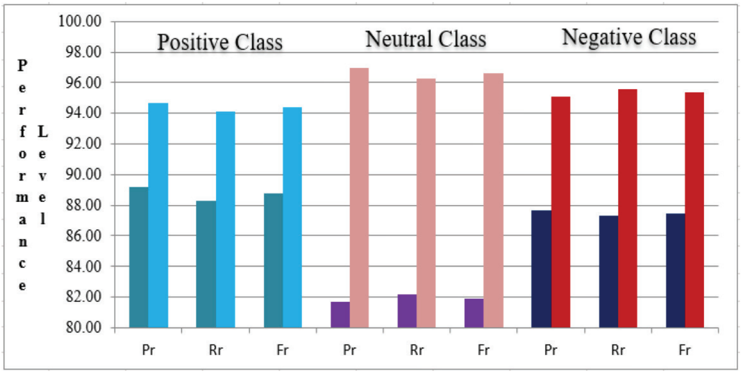


Figure 7. Performance Comparison of CNN-BiLSTM and Proposed Model

BiLSTM-Attention vs. Proposed Model: In the final phase of comparison the experiments were performed for comparison of results of BiLSTM-Attention model with proposed model (Figure 8). Comparing to the suggested technique, the CNN-BiLSTM delivers poor outcomes (*Pr*: 87.65%, *Rr*: 86.32%, *Fr*: 86.98% for positive classification, *Pr*: 93.26%, *Rr*: 93.14, *Fr*: 93.20% neutral classification, *Pr*: 90.25%, *Rr*: 91.25, *Fr*: 90.75% for negative classification and *Ar*: 92.63% accuracy). It was observed that the Attention mechanism has the disadvantage of having to pay attention to all words on the side of source for each target text, which is costly and makes translating longer sequences unfeasible (Zhang et al., 2018).

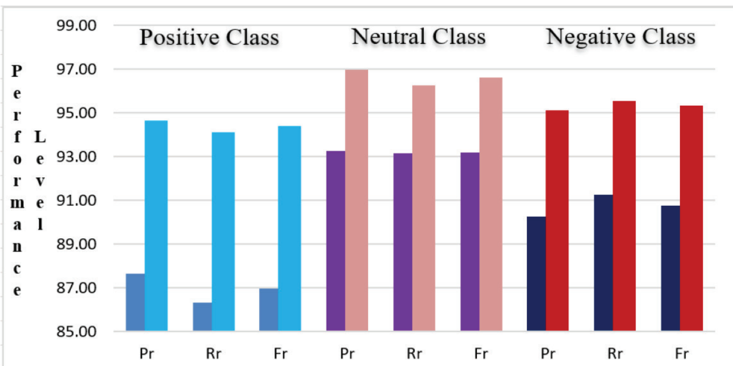


Figure 8. Performance Comparison of BiLSTM-Attention and Proposed Model

Except for the proposed model, none of the models perform consistently across all classes, as shown in Table 4. The experimental outcomes show that the BiLSTM model exhibits much lower performance for negative class classification whereas CNN-BiLSTM shows degraded performance for neutral class classification. In terms of positive class categorization, the BiLSTM-Attention model performs poorly. Finally, the consistent and highest level of performance of the proposed model was with *Pr*: 94.65%, *Rr*: 94.12%, *Fr*: 94.38% for positive classification, *Pr*: 96.98%, *Rr*: 96.25, *Fr*: 96.61% neutral classification, *Pr*: 95.12%, *Rr*: 95.56, *Fr*: 95.34% for negative classification and *Ar*: 97.01% accuracy).

Figure 9 shows the overall categorization accuracy of several deep learning approaches. The CNN approach has an accuracy of 86.35 percent, BiLSTM has an accuracy of 89.32 percent, and a CNN-BiLSTM and BiLSTM-Attention approach have an accuracy of 90.35 percent and 92.63 percent, respectively. With a 97.01 percent accuracy rating, the suggested model exceeds all previous models.

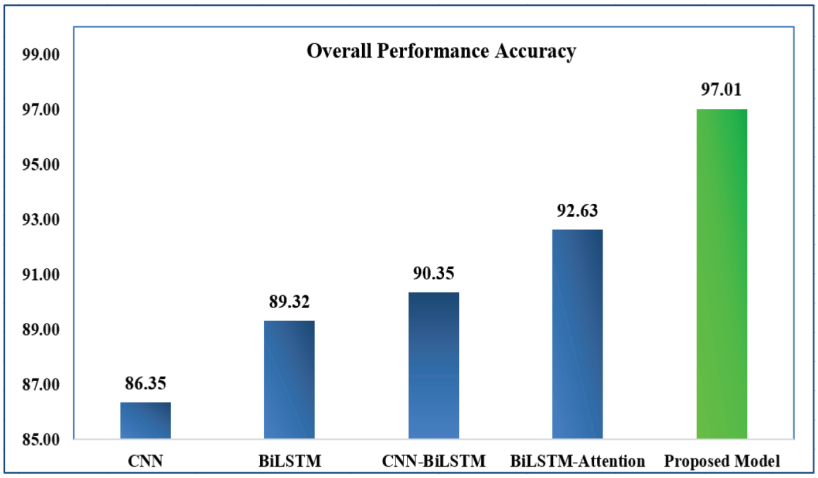


Figure 9. Performance Comparison of Overall Accuracy

5.3. Validation of Classification using ROC Curve

To verify the correctness for classification of the proposed model, a comparison of Training Accuracy and Training Loss has been performed among the deep learning model. A receiver operating characteristic (ROC) curve graph (online) has been used to illustrate the evaluation results. In Figure 10, of ROC Curve graph X-axis denotes FPR (False Positive Rate) and Y-axis denotes TPR (True Positive Rate) of the classification. The area under the curve is a number that goes from 0 to 1, the AUC (Area under the ROC Curve) under ROC values near 1 imply that the model is performing well. As observed in the figure, the performance of the proposed model is better than other models. Figure 10 shows that the proposed model shows promising results in terms of classification.

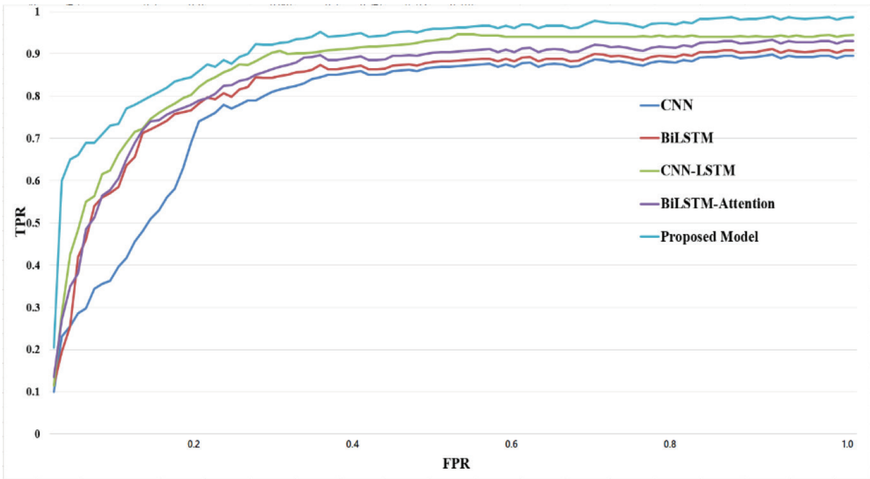


Figure 10. Comparison of the proposed model for classification using ROC Curves

In Figure 10 it is noticeable that the proposed model exhibited better performance than the other deep learning models. The observations obtained from ROC curves demonstrate that the proposed model achieves higher accuracy.

6. Conclusion

This study produced a sentiment analyzer that could be used to extract people's opinions regarding Indian Airline services as expressed on social media. The three pre-trained word embedding algorithms that were trained and assessed were Word2Vec, Glove, and FastText, with FastText outperforming the others in terms of word vectorization accuracy. Further experiments on deep learning models were implemented using FastText embeddings as it has shown the highest accuracy. The research findings revealed that the proposed model worked admirably on the obtained dataset, even beating the baseline classifier. The proposed model achieved an F-measure of 94.38% for the classification of positive reviews, 96.61% for neutral, and 95.32% for negative reviews classification. The overall accuracy achieved from the proposed model is 97.01%, which significantly outperforms other deep learning models. The results show that other models did not perform consistently for positive, negative, and neutral classes and the proposed model proved to be consistent. The proposed model achieves the highest accuracy of classification over other models. The outcomes confirmed the utility of the proposed model as a viable option for addressing the users' emotions regarding the services on social media.

References

Baziotis, C., Pelekis, N., and Doulkeridis C., 2017. Data Stories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based Sentiment Analysis. *In Proc. 11th Int. Workshop Semantic Eval* (pp. 747-754).

Chen, M., Herrera, F., and Hwang, K., 2018. Cognitive computing: Architecture, technologies and intelligent applications. *IEEE Access*, vol. 6 (pp. 19774-19783)

Chollet, F., 2017. *Deep Learning With Python*. Shelter, Island: Manning.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol* (pp. 4171-4186).

Dzikienė, K., Damaševičius, J., and Woniak, R. M., 2019 Sentiment Analysis of Lithuanian Texts Using Traditional and Deep Learning Approaches. *Computers*, *8*, 4. [Online]. Available: https://doi.org/10.3390/computers8010004

Fan, Y.X., Guo, J.F., and Lan, Y.Y., 2017. Context-based deep semantic sentence retrieval model. Chin. J. Inform. Sci. 31(5), 161-167

Gu, S., Zhang, L., Hou, Y., and Song, Y., 2018. A position-aware bidirectional attention network for aspect-level Sentiment Analysis. In *Proceedings of International Conference on Computing Linguistics* (pp. 774-784).

Guo, Y., Li, W., Jin, C., Duan, Y., and Wu, S., 2018. An integrated neural model for sentence classification. In *Proc. Chin. Control Decis. Conf. (CCDC)* (pp. 6268-6273).

Han, H., Bai, X., and Li, P., 2019. Augmented sentiment representation by learning context information. *Neural Comput. Appl.*, vol. 31, n. 12 (pp. 8475-8482).

Hu, R., Rui, L., Zeng, P., Chen, L., and Fan, X., 2018. Text Sentiment Analysis: A review. *In Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2018 (pp. 2283-2288).

Huang, B., Qu, Y., and Carley, K., 2018. Aspect level sentiment classification with attention-over-attention neural networks. In *Proc. Conf. Social Comput* (pp. 197-206).

Huang, Z., Xu, W., and Yu, K., 2015. Bidirectional LSTM-CRF models for sequence tagging. *arXiv: 1508.01991*. [Online]. Available: https://doi.org/10.48550/arXiv.1508.01991

Hwang, K., and Chen, M., 2017. Big Data Analytics for Cloud. *IoT and Cognitive Computing*. Hoboken, NJ, USA: Wiley,

Jiang, Z., Gao, S., and Chen, L., 2019. Study on text representation method based on deep learning and topic information. *Computing*, vol. 102, n. 3 (pp. 623-642).

Kim, K., Chung, B. S., Choi, Y., Lee, S., Jung, J.-Y., and Park, J., 2014. Language independent semantic kernels for short-text classification, *Expert Syst. Appl.*, vol. 41, n. 2, pp. 735_743, Feb. 2014.

Lebret, R., and Collobert, R., 2013. Word emdeddings through Hellinger PCA., arXiv:1312.5542. [Online]. *Available:* https://doi.org/10.48550/arXiv.1312.5542

Liu, B., 2012 Sentiment Analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol*., 5, 1-167.

Liu, G., Xu, X., Deng, B., Chen, S., and Li, L., 2016. A hybrid method for bilingual text sentiment classification based on deep learning. In *Proc. 17th IEEE/ACIS Int. Conf. Softw. Eng*., *Artif. Intell*., *Netw. Parallel/Distrib.Comput. (SNPD)* (pp. 93-98).

Liu, W., Cao, G., and Yin, J., 2019. Bi-level attention model for Sentiment Analysis of short texts. *IEEE Access*, vol. 7, (pp. 119813-119822).

Ma, Y., Peng, H., Khan, T., Cambria, E., and Hussain, A., 2018. Sentic LSTM: A hybrid network for targeted aspect-based Sentiment Analysis. *Cognitive Computing*, vol. 10, n. 4.

Ma, X., and Hovy, E., 2016. End-to-end sequence labeling via bi-directional LSTM-CNNS-CRF. *arXiv:1603.01354*. [Online]. Available: https://arxiv.org/abs/1603.01354

Neethu, M. S., and Rajasree, R., 2013. Sentiment Analysis in Twitter using machine learning techniques. 2013, *In Proc. 4th Int. Conf. Comput*., *Commun.Netw. Technol. (ICCCNT)*, pp. 1-5.

Pang, B., and Lee, L., 2008. Opinion mining and Sentiment Analysis. *Found. Trends Inf. Retr*., 2, 1-135.

Pennington, J., Socher, R., and Manning, C., 2014. Glove: Global vectors forword representation. In *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)* (pp. 1532-1543).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer L., 2018. Deep contextualized word representations. *Journal of Associative Computing Linguistics*, vol. 1 (pp. 2227-2237). [Online]. Available: https://www.quora.com/What-are-some-of-thelimitations-or-drawbacks-of-Convolutional-Neural-Networks

Pham, D.H., and Le, A.C., 2018. Exploiting multiple word embeddings and one-hot character vectors for aspect-based Sentiment Analysis. *Int. J. Approx. Reasoning*, vol. 103 (pp. 1-10).

Rehman, A. U., Malik, A. K., Raza, B., and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews Sentiment Analysis, *Multimedia Tools Appl.*, vol. 78, n. 18 (pp. 26597-26613).

Rezaeinia, S. M., Rahmani, R., Ghodsi, A., and Veisi H., 2019. Sentiment Analysis based on improved pre-trained word embeddings. *Expert Syst. Appl.*, vol. 117 (pp. 139-147)

Seo, S., Kim, C., Kim, H., Mo, K., and Kang, P., 2020. Comparative Study of Deep Learning-Based Sentiment Classification. In *IEEE Access*, vol. 8 (pp. 6861-6875).

Simonyan, K., and Zisserman, A., 2014. Two-Stream Convolutional Networks for Action Recognition in Videos. *London, U.K.: Univ. of Oxford*

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment Treebank. *In Proc. Conf. Empirical Methods Natural Lang. Process* (pp. 1631-1642).

Sun, B., Tian, F., and Liang, L., 2018. Tibetan micro-blog Sentiment Analysis based on mixed deep learning. In *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)* (pp. 109-112).

Tang, D., Qin, B., Feng, X., and Liu, T., 2016.Effective LSTMs for target- dependent sentiment classification. In *Proc. COLING 26th Int. Conf. Comput. Linguistics* (pp. 3298-3307).

Tay, Y., Tuan, L., and Hui, S., 2018. Learning to attend via word-aspect associative fusion for aspect-based Sentiment Analysis. In *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)* (pp. 5956-5963).

Wint, Z. Z., Manabe, Y., and Aritsugi, M., 2018. Deep learning based sentiment classification in social network services datasets. In *Proc. IEEE Int. Conf. Big Data, Cloud Comput*., *Data Sci. Eng. (BCD)* (pp. 91-96).

Xu, G., Meng, Y., Qiu, X., Yu, Z., and Wu, X., 2019. Sentiment Analysis of comment texts based on BiLSTM,. IEEE Access, vol. 7, pp. 51522-51532.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E., 2016. Hierarchical attention networks for document classification. *In Proc Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol* (pp. 1480-1489).

Yao, X. L., 2017. Attention-based BiLSTM neural networks for sentiment classification of short texts. In *Proceedings of International Conference Information Science Cloud Computing* (pp. 110-117).

Yin, W., Kann, K., Yu, M., and Schütze, H., 2017. Comparative study of CNN and RNN for natural language processing, 2017, arXiv: 1702.01923. [Online]. Available: https://doi.org/10.48550/arXiv.1702.01923

Yu, Q., Zhao, H., and Wang Z., 2019. Attention-based bidirectional gated recurrent unit neural networks for Sentiment Analysis. In *Proc. 2nd Int. Conf. Artif. Intell. Pattern Recognit.* Cham, Switzerland (pp. 67-78). Springer.

Zhang, Y., Wang, J., and Zhang, X., 2018. YNU-HPCC at SemEval-2018 task 1: BiLSTM with attention based Sentiment Analysis for affect in tweets. In *Proceedings of 12th Int. Workshop Semantic Eval.* (pp. 273-278). [Online]. Available: https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

Zheng, J., and Zheng, L., 2019. A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification, *IEEE Access*, vol. 7 (pp. 106673-106685)

Zhou, K., and Long, F., 2018. Sentiment Analysis of text based on CNN and bi-directional LSTM model. In *Proc. 24th Int. Conf. Autom. Comput. (ICAC)*, (pp. 1-5).

Zhou, X., Wan, X., and Xiao, J., 2016. Attention-based LSTM network for cross-lingual sentiment classification. In *Proc. Conf. Empirical Methods Natural Lang. Processing* (pp. 247-256).

Zhu, Y., Gao, X., Zhang, W., Liu, S., and Zhang, Y., 2018. A bi-directional LSTM CNN model with attention for aspect-level text classification. *Future Internet*, vol. 10, n. 12,116.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 331-348
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.28331

Artificial Intelligence (AI) in Advertising: Understanding and Schematizing the Behaviors of Social Media Users

Metin Argan[a], Halime Dinc[b], Sabri Kaya[c] and Mehpare Tokay Argan[d]

[a] Faculty of Sports Science, Eskisehir Technical University, Eskisehir, Turkey
[b] Faculty of Sports Sciences, Afyon Kocatepe University, Afyon, Turkey
[c] Faculty of Sports Science, Kirikkale University, Kirikkale, Turkey
[d] Faculty of Applied Sciences, Bilecik Seyh Edebali University, Bilecik, Turkey
✉ margan@eskisehir.edu.tr, halimedinc@aku.edu.tr, sbrkaya@gmail.com, mehpare.argan@bilecik.edu.tr

KEYWORDS

artificial intelligence; advertising; social media user; AI ads effects

ABSTRACT

Nowadays, information technology is not only widely used in all walks of life but also fully applied in the marketing and advertisement sector. In particular, Artificial Intelligence (AI) has received growing attention worldwide because of its impact on advertising. However, it remains unclear how social media users react to AI advertisements. The purpose of this study is to examine the behavior of social media users towards AI-based advertisements. This study used a qualitative method, including a semi-structured interview. A total of 23 semi-structured interviews were conducted with social media users aged 18 and over, using a purposive sampling method. The interviews lasted between 27.05–50.39 minutes on average (Mean: 37.48 SD: 6.25) between August and October 2021. We categorized the findings of the current qualitative research into three main process themes: I) reception; II) diving; and III) break-point. While 'reception' covers positive and negative sub-themes, 'diving' includes three themes: comparison, timesaving, and leaping. The final theme, 'break-point', represents the decision-making stage and includes negative or positive opinions. This study provides content producers, social media practitioners, marketing managers, advertising industry, AI researchers, and academics with many insights into AI advertising.

1. Introduction

Due to the advancement of internet technology, big data, computing power, sensor communication (Ashraf et al., 2020a), and algorithmic breakthroughs, the impact of robotics and artificial intelligence technologies on human life is becoming ever more clear. Artificial intelligence (AI), which is a branch of information technology innovation, is a key factor for business and academics. Robotics and artificial intelligence (AI) have been among the main drivers of almost every field in the last decade. AI provides models of human intelligence for cognitive psychology, as well as computer software, hardware, and robotics that perform intelligent human functions (Churcher, 1991). The concept of AI, in a broad sense, refers to a set of machine functions that can learn with the help of humans or entirely on their own (Rodgers, 2021). According to another, more comprehensive definition «AI is the ability for a machine to collect information and use-sophisticated algorithms and logical functions to learn from it, there-by adapting future capabilities based on additional information to increase knowledge» (McCarthy and Hayes, 1981; Xian, 2021).

Recent technological advances, especially those leveraging artificial intelligence (AI) and machine learning are challenging the concept of marketing, contemporary advertising, and advertising content (Campbell et al., 2022). It has been widely applied Ai at in the advertising industry and increasingly attracts the attention of marketing scholars (Wu et al., 2021). While some other technologies may perform human-specific functions, AI also aims to achieve unique human characteristics such as speech, vision, reasoning, planning, and creativity. AI is rapidly becoming more central to the day-to-day digital world, and the marketing and advertising world is no exception (Murgai, 2018). Considering the technological advances in advertising, perhaps none are as exciting as AI (Rodgers, 2021).

Current developments and figures also point to the importance of AI in marketing and the high probability of this synergy continuing to grow in the future. Over 75% of consumers already use an AI-powered service or device. An expected 53% growth is expected in AI marketing in 2021. By 2023, we expected global digital advertising to reach $517.51 billion, with AI accounting for 80% of this sum (Rodgers, 2021). As AI quickly becomes more sophisticated and widely adopted in marketing, the ability of marketers to effectively implement and manage it will become an ever more important skill (Shah et al., 2020). As in almost every field, AI has created unique opportunities to deliver personalized advertising messages to consumers. The reasoning capabilities of AI mean it can reveal personality, tendencies, values, and needs from social media users' comments and posts (Kietzmann et al., 2018).

Research into online shopping and product research is concerned with achieving customer satisfaction; it is in the interests of brands to leverage the capabilities of AI algorithms for personalized product recommendations that meet the needs of their target audiences. AI therefore creates a great opportunity for advertisers to target audience through personalized messages. With the proliferation of AI use, scientists have turned their attention to the effects of such technologies in personalized advertising. The best way to understand such effects would form a consumer/user-oriented perspective.

Based on consumers' searches on the internet or social media, AI algorithms aim to produce creative solutions for the products or services in which the user may be interested. Although the personalized messages can have a high accuracy rate sometimes, it may be the opposite in others where there is no potential consumption of the product. For example, for a low-income teenager who is searching for BMW cars out of curiosity, the effectiveness of such a category is questionable. Therefore, it can be said that AI advertising messages should be based on new algorithms supported by various parameters, such as the potential consumer's lifestyle and annual income.

Some of the previous research on this subject had focused on examining the wide range of tasks of multi criteria assessment of customer service in various social fields —retail stores, medicine, culture, health, physical educational training, public catering, other household and domestic services (Ashraf et al., 2020b)— and on different perspectives on AI technology (Zhao et al., 2020).

Studies that reveal the use of AI in the con-text of leisure services seem to focus only on subjects such as advantages and disadvantages, as well as factors that lead to the adoption of technologies. Although there are some studies such as using AI predicted personality to enhance advertising effectiveness (Shumanov et al., 2021), AI in advertising creativity (Vakratsas and Wang, 2020), promises and perils of AI (Rodgers, 2021), Twitter conversations about AI in advertising (Wu et al., 2021), the number of studies that address the effectiveness of AI ads from a consumer-oriented and comprehensive perspective, is low.

Unlike other studies, in this study, consumers' behaviors/reactions to AI advertisements are discussed by using a qualitative research method. This qualitative research focuses on consumers' reactions to AI-based ads for leisure services. Therefore, in qualitative consumer reactions to both leisure services and AI-ads, this study fills a significant gap in the literature. As a result, based on prior evidence, we believe that studying AI in leisure advertising can contribute to a better identification of the parameters that potentially impact consumers.

2. Literature Review

2.1. AI in leisure

Technology has blurred the lines between work and leisure (Lashua, 2014, 2018), which have had its own advantages and disadvantages. It shows that AI has become a guiding force. For example, the opinions based on a Washington Post report revealed that AI would have a negative impact on employment, although it would give the industry an advantage because of the opportunities offered by digitalization and algorithms. Studies have emphasized that AI can avoid monotonous, «boring» work and to increase the creativity of employees. Some facilitating factors related to AI, such as time saving, improved service, eliminating boredom through variation and creativity can be associated with leisure (Churcher, 1991).

For example, Hou (2013) underlines that forecasting traffic flow at leisure farms could be done using the advanced AI technology of artificial neural networks (ANNs), thus it would contribute positively to users. According to the research conducted by Xian (2021) regarding the adoption of AI technology in the leisure economy, seven factors were predicted to be effective performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, habit, and personal innovativeness. The performance expectation underlines the degree to which leisure activities increase efficiency because of using AI. Effort expectancy represents the user's expectation of convenience achieved through AI.

Consumers should feel that using this technology is easy. Social influence represents the influence of friends, relatives, family and peers on using this technology. The social media user can learn to use this technology through the influence of the social environment in which they interact. Facilitating conditions represent all kinds of factors that aid in the consumer's acceptance of AI-related technologies, such as online instruction and ongoing support. Hedonic motivation reveals the joy and pleasure derived from using technology. The state of not being bored or having fun is quite important, especially with Millennial and Generation Z. Price value is related to the cost of using this technology but can also be associated with variables such as the duration of the search for the product, and the savings provided by the offered alternative. Habit is about positive outcomes from experiences. If the consumer has had a positive experience using AI, they are more likely to continue using it.

Finally, personal innovativeness relates to the consumer's preference, risk-taking, or predisposition to use innovations. Some consumers differ from others in adopting new products. This situation is considered as adopting innovations. The higher the consumers ' expectations of the AI service are, the more likely they are to use it. Because of the research, it was found that the remaining six factors, excluding effort expectancy, influenced acceptance intention (Xian, 2021).

2.2. AI Advertising and Consumer Behavior

Advertising media is undergoing major changes. The online media advertising market is expanding. These changes significantly affected advertisements in traditional media. In various researches, they have addressed different aspects of the ad management process regarding the use of AI in advertising; such as market research, targeting and media selection, ad creation and design, ad placement and execution, ad performance and purchase decision, machine learning (Rodgers and Nguyen, 2022; Shah et al., 2020).

The study on the promises and perils of AI advertising (Rodgers, 2021) considered six articles, and the advantages and disadvantages of each subject were emphasized. The first of these articles looked at AI Influencers as brand endorsers. The second article examined how ad placement effects can be improved using machine-learning algorithms. The third article was on creativity in advertising and the fourth one addressed image-text mismatch. The fifth article was about the stimulating or intimidating effect, and the last article regarded a text analysis program that revealed linguistic questioning and word calculation. All these elements underline the need to evaluate multiple approaches in order to get effective consumer-oriented results in the AI advertising approach.

In the study on the use of machine learning in advertising, three approaches to the target setting were emphasized. These are user-centric, content-centric and click fraud approaches. The user-centric approach includes behavioral targeting and user profiling. The content-centric approach covers elements such as real time bidding and contextual advertising (display, video, vehicle, blog, web doc) (Choi and Lim, 2020). Click fraud is a challenging topic in AI advertising, because it misjudges behavior and incurs additional costs.

There are several steps to consider regarding the influence of AI on consumer experience. The first of these is the recognition of wants and needs. Attempts have been made to define the characteristics of smart advertising in the literature through the analysis of AI-powered applications. The crucial point to achieve so, is the personalization of intelligent ads according to consumers' specific requirements, needs & interests, and life-style. Going beyond mere guesses of the user's interests, intelligent ads can accurately predict the user's wants and needs in various contexts and specific time periods, and can suggest user-specific offers (Li et al., 2002). Hence, knowing the psychology of the consumer or social media users becomes a necessity.

Emotionally appealing advertisements are meant to appeal to the consumers’ heart, making them feel special, and a part of the brand (Mogaji, 2018). It primarily served emotional ads to elicit positive feelings and emotions (Aaker and Stayman, 1992). Using Microsoft's AI system Azure for consumer profiling, the system analyzes billions of data points in seconds to identify the needs of individuals. It then personalizes the content in real time to align with those consumer interests. As user interactions accumulate, more personal data makes it possible to optimize smart ads and serve the user better, thus establishing a long-term brand relationship (Li, 2019; Lai, 2020). As the consumers' digital footprints evolve through social-media status updates, purchasing behavior, or online comments and posts—machine learning continuously updates these profiles. AI also helps advertisers' «manifest» consumers' needs or wants. Similarly, Pinterest employs image recognition to learn about individual users' particular style preferences through the images they have pinned on the site. The website then suggests other relevant images that reflect the user's specific preferences, thus facilitating need or want recognition (Kietzmann et al., 2018).

A concept that may be related to the first step is FoMO (fear of missing out). FoMO is a «pervasive apprehension that others might have a rewarding experience from which one is absent, FoMO characterized by the desire to stay continually connected with what others are doing» (Przybylski et al., 2013). FoMO has successfully used in commercial advertising calls to start sales (Argan et al., 2018). For example, messages about «an opportunity not to be missed» can be effective in many hotels or leisure services. If some individuals are susceptible, i.e., reactive to fear appeals, FoMO can be used as a strategic tool. The effectiveness of FoMO appeals tested by introducing advertisements and stimulus materials. According to research on the effectiveness of FoMO, respondents reported commercial FoMO appeals started personally by «salespeople» or impersonally through «advertisements» and «sales catalogs (Hodkinson, 2019). There are many advertisements in practice showing that FoMO is widely used for the success of travel promotions. Some examples of advertisements (Hodkinson 2019), put forward by citing the industries were «Winning with FoMO', «Cruiselings: A New Breed of FoMO» and 'MTV and Flight Center Create Travel «FoMO».

Similarly, some observe that FoMO is used directly or indirectly in the content of many advertisements (JWT, 2012). It has also shown that the FoMO strategy is used to emphasize the sense of scarcity in advertising messages (Hodkinson, 2019). In particular, the messages in the FoMO advertising campaign can impact young audiences. Used in the first evaluation, which is another step; Google AdWords lets marketers make better distinctions to target unqualified and qualified leads better. With Artificial Intelligence, Google analyzes search query data by analyzing not just keywords, but also context words and phrases and other big data. From that point, Google determines potentially useful consumer subsets and more accurate targeting. With Artificial Intelligence, wants and needs can be understood in real-time as customers communicate them digitally and «richer» profiles can be built quicker. AI also allows advertisers to manifest the wants and needs of individuals and increase the quality and quantity of their sales. (Kietzmann et al., 2018).

According to the third step, the active evaluation, advertising aims to persuade consumers to trust the offer and convince them that they are making the best choices when they narrow down their list of brand preferences (Batra and Keller, 2016). One strategy is to target high-intent consumers and provide them with much persuasive content. Since trust is the building block in intelligent advertising, the consumer should not be coerced into content, but instead, the content should be shown to the user with their consent (Li et al., 2002). Machine learning and AI techniques help marketers narrow down targeted customers through digital advertising and generate efficient results (Shah et al., 2020).

Previously, most advertising messages sent to unconstrained customers were going to waste, which increased cost and reduced the effectiveness. For example, instead of randomly recommending new movies to a moviegoer, it would be more effective for the algorithm to recommend a new adventure movie based on the few adventure movies they have watched before. It emphasized three points in active evaluation. According to the first point, predictive lead scoring through machine learning. The algorithm collects verified existing customer data and recognizes trends and patterns; and then, after being fed with additional external data about consumer activities and interests, it creates reliable lead profiles. The second point focuses on machine learning and enabling the editing of advertising content by learning from consumer behavior in real time through functions such as image recognition, speech recognition and natural language processing (NLP) (Kietzmann et al., 2018). The third and final point is about emotion AI. Marketers use emotion AI to pretest ads and understand what consumers are saying and feeling based on publicly available data, such as reviews, blogs or videos about brands (Kietzmann et al., 2018).

In the fourth step, which is related to purchasing, advertising aims to take consumers out of the decision journey and take them into action while they decide on the value of the brand and how much they want to pay (Batra and Keller, 2016). Advertisers' high-light information about warranties, returns, purchase incentives, amenities, and places to buy. In the last stage, post-purchase, consumers evaluate their satisfaction and, perhaps by word of mouth, they discuss whether they want to buy the product again (Kietzmann et al., 2018). Advertisers try to please them by reinforcing satisfaction and eliminating potential problems. I enabled «chat-bots» can interact with customers. It can use parameters such as determining the most valuable customer, calculating customer lifetime value, likelihood of re-engagement and the probability of leaving (Kietzmann and Canhoto, 2013).

3. Method

3.1. Procedures

It followed a method of collecting qualitative data through photo-elicitation to investigate the AI advertising experiences and reactions of consumers. It is possible to better understand AI ad experiences in terms of data quality by talking about multiple realities with people who have been exposed to AI ads or have experienced AI in terms of leisure services (travel, entertainment, vacation, etc.). For this reason, it included people who have experienced AI ads in the research as a prerequisite. Interviews comprised two stages, a regular interview as a conversation and a photo-elicitation.

The interviewees were asked to show and describe one or more images that best described their AI advertising experience. Images were mostly used to make it easier for social media users/consumers to describe their concrete experiences. Using images was intended to provide an entertaining and engaging dialog between the interviewer and the interviewees. Questions from the personal information form (such as gender, age, online shopping frequency, and social media behavior) were asked of all participants and filled in by the second author. Interviews were conducted via videoconference (with 8 participants via Zoom), and face-to-face (with 15 participants).

The interviews lasted between 27.05–50.39 minutes on average (Mean: 37.48 SD: 6.25) and were conducted with participants by the second author between August and October 2021. All interviews were digitally recorded and transcribed. Since the primary aim of the study was to explore how consumers react to AI advertising in a broad sense, we asked open-ended questions such as 'What do you think about AI and intelligent advertising?' or 'What is your best experience with AI advertising so far?' Other central questions to the study included: 'What do you think about the AI advertising you receive about leisure services?', 'Did you open and review these ads? What did you do? Why?', 'While you are searching for any product (for example, holiday, concert, ticket, travel, accommodation, etc.) on the Internet or social media, do you deliberately leave a trace to see other opportunities (affordable price, variety of quality, etc.)?', 'In which cases do you click and open AI ads?', and 'what does this experience mean to you?'.

3.2. Sampling

Participants were selected from people who reported having experienced AI ads. The strength of the sample was based on specific facts related to the respondents' knowledge, experiences, or properties on the subject under investigation (Maltured et al., 2016). A total of 23 interviews were conducted with social media users aged 18 and over, using a purposive sampling method. When determining the sample size in this study, researchers applied saturation as a guiding principle during data collection (Glaser and Strauss, 2017). As the freshly collected data did not create new insights on the subject, it stopped data collection after 23 people in this study. Of the participants, 52.2% were female and 47.8% were male. Participants comprised young individuals between the ages of 19-23 (21.04±1.06). In terms of income level, 34.8% (8 people) had a monthly income of between 5001 TL to 7500 TL. Regarding social media usage behavior, 34.8% (8 people) of the participants had been using social media for at least 6 years and 52.2% (12 people) reported that they were active on their social media accounts for 2-4 hours a day. The most frequently searched leisure service was a concert event, which is 39.1% of the participants (9 people). Additionally, 34.8% (8 people) of the participants stated their frequency of online shopping as «very often». It showed other information on the characteristics of the participants in Table 1.

Table 1. Profile of participants

No.	Pseudonym	Gender	Age	Income (TL)	Social media usage time	Social media usage year	Leisure services	Online shopping frequency
1	Kerem	M	21	7501-10000	2-4 hours	11,00	Travel Ticket	Often
2	Aysun	F	21	7501-10000	More than 6 hours	7,00	Concert	Frequent
3	Deniz	M	19	5001-7500	2-4 hours	6,00	Concert	Moderate
4	Asuman	F	21	5001-7500	2-4 hours	9,00	Concert	Often
5	Can	M	22	7501-10000	2-4 hours	8,00	Concert	Rare
6	Melahat	F	22	7501-10000	2-4 hours	8,00	Concert	Rare
7	Hilal	F	22	5001-7500	4-6 hours	6,00	Vacation	Moderate
8	Rıdvan	M	21	10000 and >	2-4 hours	5,00	Travel Ticket	Often
9	Oğuz	M	21	5001-7500	Less than 2 hours	10,00	Travel Ticket	Frequent
10	Büşra	F	23	7501-10000	Less than 2 hours	10,00	Travel Ticket	Frequent
11	Sude	F	20	5001-7500	More than 6 hours	6,00	Concert	Often
12	Ekrem	M	21	2501-5000	Less than 2 hours	12,00	Another Camp	Frequent

Table 1. Profile of participants (continued)

13	Yusuf	M	22	2501-5000	Less than 2 hours	5,00	Vacation	Often
14	Mine	F	23	2500 <	2-4 hours	6,00	Cinema	Often
15	Leman	F	21	2501-5000	2-4 hours	4,00	Vacation	Moderate
16	Ismail	M	21	5001-7500	2-4 hours	4,00	Vacation	Rare
17	Müge	F	20	2500 <	Less than 2 hours	6,00	Vacation	Moderate
18	Selçuk	M	20	2501-5000	2-4 hours	7,00	Concert	Moderate
19	Seda	F	22	5001-7500	More than 6 hours	8,00	Concert	Frequent
20	Eylül	F	21	5001-7500	Less than 2 hours	6,00	Vacation	Often
21	Ebru	F	21	2501-5000	2-4 hours	8,00	Vacation	Often
22	Veli	M	20	7501-10000	2-4 hours	6,00	Concert	Frequent
23	Okan	M	19	2501-5000	4-6 hours	6,00	Vacation	Rare

3.3. Data analysis

As suggested by Charmaz (2006) and many qualitative studies, data analysis, data collection, and note-taking were conducted simultaneously. The interviews recorded at the end of the data collection process were read, coded, and categorized as themes, topics, and sub-topics. Qualitative data analysis procedures, as suggested by Braun and Clarke (2006), apply to this study. The recommended steps, according to these procedures, are (1) data familiarity, (2) creating an original technique for coding responses, (3) developing a method by which themes are discovered, (4) developing evaluation techniques for themes, (5) determining definitions and choosing names of all themes, and (6) compiling and writing the final report. Four credibility assessments were used throughout the process (Creswell, 2013). First, the results of the analysis were confirmed by making supportive observations during the interviews. Second, the consensus among researchers was established, as suggested in many qualitative studies. In the third step, which is expert opinion, a PhD-level expert on advertising confirmed the themes and categories' suitability for the research efforts. Fourth and lastly, a member-checking approach was used (Table 2). The results were shown to two interviewees and confirmed by them with minor corrections.

Table 2. Example of coding

Meaning unit	**Initial coding**	**Sub-theme**	**Theme**	**Memo**
Ekrem: Time is precious for every person. It's the same in me. When I review the content, I think my time was completely wasted.	The participant felt that his time was wasted while reviewing the content of some suggested ads.	Time-saving	Diving-in	When people review ads, they look for speed in getting the results they want. If the ad causes the person to think of wasting unnecessary time, the ad review is terminated.

4. Results

Because of the interviews with the participants, it was found that the advertising process was experienced in 3 stages. The researcher evaluated as «process themes» these stages, which emerged spontaneously during the interviews. The responses given within the resulting «process themes» were treated as «topics». The process themes were named: 1. «reception», 2. «diving-in», 3. «breakpoint». It addressed the responses of the participants at the above stages within the «topics» from which relevant themes spontaneously emerged. At the initial stage (reception), it classified the responses of the participants as positive and negative receptions, and the information obtained from the participants' views on the factors that led to positive and negative processed reactions under «sub-topics». The second phase, which emerges after the advertisement successfully passed the «reception» phase, was named «diving-in», and the reactions and expectations at this stage were also discussed. Sub-topics were obtained from the topics that emerged because of the interviews, and it compiled the responses for the 2nd stage. Finally, the phase of ending the advertisement (Breakpoint) was reached with the guidance of the factors affecting the success of the «diving-in» process in terms of participant response. At this stage, the main factors that cause the termination of the ads and both their positive and negative effects on the participant's reaction were discussed. It gave the schema that emerged as a result of the study in Figure 1, and we explained it with quotations of the views expressed by the participants.

4.1. Reception

The ad sent by AI either reaches the ad involvement or is skipped without being viewed, depending on the reaction of the user when it first appears. The inferences made because of the interviews show that, at this stage, positive and negative receptions govern the process. While general factors such as attracting the attention of the person and creating curiosity can be counted among the factors that affect negative reception, the opposite of the explanation, given in the «sub-topics» of the topic, «positive reception» can be counted as negative reception. I mention more specific cases in the negative reception topic below.

4.1.1. Negative Reception

When advertisements related to leisure services reach the consumer, the image first encountered by the person determines the perception of the advertisement. Considering the average age of the sample group, it is likely that they often consider their financial interests. Therefore, while visual attractiveness is a serious factor, the perception of expensiveness can hinder this process. It gave an opinion of a participant in favor of this inference below: «I skip the ad when I see that a place or a hotel seems luxurious, thinking that I can't afford it, anyway. I have to consider my wallet as much as I would like to be in such places» (Oğuz, 21).

Another negative reception factor is the traces left by experiences. Participants avoid re-experiencing prior experiences that did not satisfy them and generalize similar activities. It gave the thought, reflecting this view below: «…last year we went to Kaş. My foot was seriously injured by the cliffs and my vacation was ruined. Whenever I get an advertisement about Kaş, I don't even look at it» (Sude, 20).

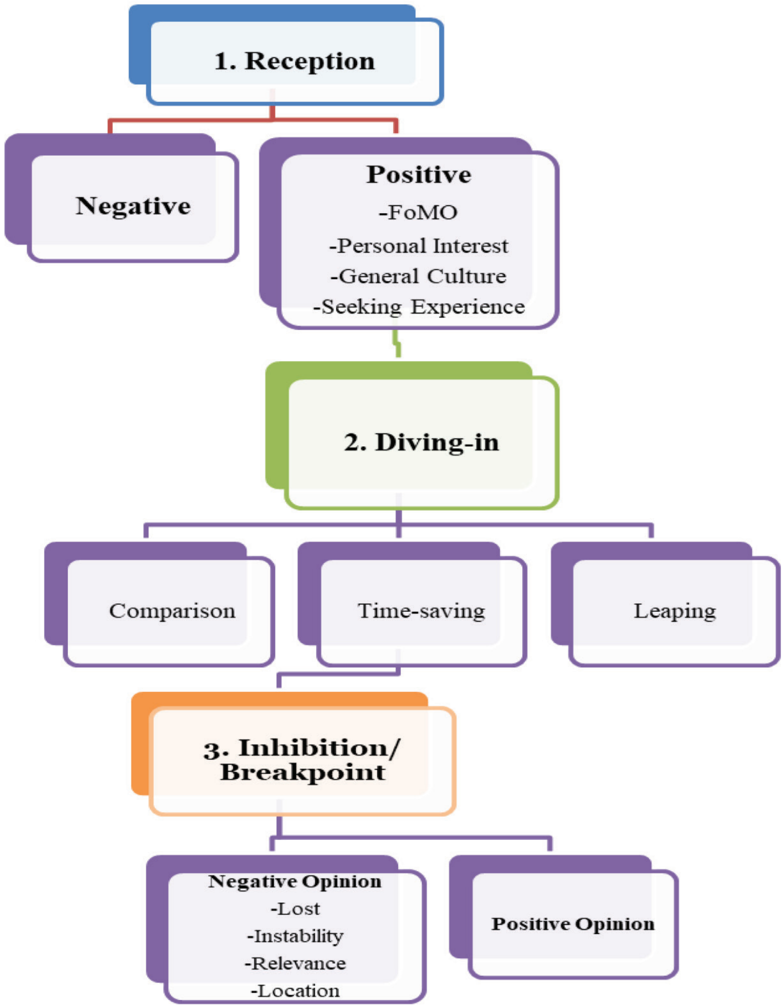


Figure 1. Consumer Reaction Towards AI Ads

4.1.2. Positive Reception

The topic covered in this subsection related to the positive receptions of individuals at the time of their first interaction with the ads. If individuals encounter positive reception elements, they can decide to continue viewing the ads. Ad viewing ends and the advertising process cannot continue if consumers cannot relate to the concepts included in positive reception. The elements of the positive reception sub-topic are FoMO, Personal Interest, General Culture, and Experience Seeking. These elements are explained below and the participants' comments are provided.

FoMO: When the fear of missing out is imposed in advertisements containing leisure services, this fear stimulated the person and opens the advert and examine it. While this may be an opportunity offer, it may also be limited-day offers and financial opportunities. Particularly, in line with the financial possibilities of the sample group, the participants state that specifying the price of the event on the main image may be an important factor in opening the relevant ad. The participants' comments supporting this view are given below.

> I usually check to see if there is a fee on the poster. Even if it's not on the poster, I'll look at the description. I mean, you should be able to just click and see it. There are already a lot of ads, so I will not bother attempting to learn the price. (Selcuk, 20)

Personal Interest: Leisure service ads that individuals encounter regarding their field of interest can attract their attention at first sight. Advertisements about the concert of an artist that the individual is a fan of, or advertisements containing a vacation destination that the individual is planning to go to, can attract him/her personally because the content includes their interests. Showing the importance of personal attention, one participant who received the advertisement through a friend commented as follows.

> Melek Mosso was coming to Afyon. I actually like her a lot, but I didn't get any ads about it. My friend randomly got the ad when I was with her. We bought our ticket as soon as we saw it. I would've missed the concert if I didn't see the ad. (Aysun, 21)

General Knowledge: Considering the above statement from another perspective, we can say that advertisements support the situations of following and attracting attention, as well as obtaining information, learning, and being aware. Rather than an opportunity situation, the content that appears in front of the user in recommended ads, which they have not searched about before and which allows him/her to get new information the elements of «being aware and curiosity» are seen as other factors that increase the chance of the ads being viewed. Besides this opinion, other participant comments are:

> I mean, sometimes these ads lead me to learn new things. A turned out there is a place called «The Hobbit Houses» in Sapanca. Where I've come across a few times and even though it was a movie scene or something. It turned out to be in some park in Turkey. I learned about it through the ads. (Deniz, 19)

> I watch a lot of movies and sometimes it’s hard to find something to watch. I've watched many movies thanks to the suggestions. (Okan, 19)

Seeking New Experiences: Advertisements containing activities that are new to the user and which they think they could experience may also be effective if the content can respond to their wants and needs. The users can have a positive attitude towards the content, which contains activities that excite them and that they are open to trying. When people encounter such content, they imagine themselves in the event and may continue their advertising involvement with the belief that they can make their wish come true. The participant's comment for the sub-topic in question is as follows:

> There are some ads that only have images of hotels and seascapes, which are already what you expect from a holiday. There was this ad that showed young people having fun on a banana boat. That's, for example, something I would like to take part in. (Kerem, 21)

4.2. Diving-in

The second theme for advertisements on leisure services is called diving. After the consumers react to the reception stage positively, they give new reactions in the review stage. The diving theme includes an ongoing process of individuals' perceptions of the ads. Consumers who cannot last in this phase do not perform their purchasing behavior.

The nature of the situations encountered during the evaluation of the contents comes to the fore. The generalization tendency of the consumer emerges because of the frequency of the encountered phenomenon and the co-existence of similar results. Content adequacy/insufficiency manifests itself as a re-reaction in the reception phase for the next ad. The reactions encountered during the diving process are associated with the sub-topics referred to as «comparison», «general knowledge», «time-saving» and «leaping». It also gives the relevant opinions below.

4.2.1. Comparison

Individuals can find a serious comparison opportunity within the suggestions given without the need to conduct additional research on the advertisement they are interested in. Digital traces gain importance here and offer users the opportunity to encounter other trace-based benchmark products later on. Users know that they will encounter these suggestions before they make a purchase, and they try to use this situation to their ad-vantage. In this process, encountering similar ads related to purchasing behavior that has not yet been performed can have a significant impact on the reminder feature and turn the intention into action. It gave opinions on participants reflecting on this topic below.

> I talked to my friends about whether we should go to the camp, and then the advertisements about the camp appeared immediately. We came across a lot of company advertisements that organize camps. Since it coincided with the conversation, we looked into them. We looked at the pages of the companies in the recommended ads, looked at their followers, their activities, prices, etc. We made a shortlist to choose from.…. Then, when I encountered ads again, I started sending them to my friends. Although the conversation was on the table, we couldn't make any real plans. But thanks to the pages I've looked through again and again; we finally went there. (Yusuf, 22)

> My parents are on vacation right now. Let me tell you how it happened. They didn't have a vacation plan that I knew of. Ads kept popping up in front of my mom. She was saying things like «maybe we should go» or something, but no conclusion. Again, while I was at home like this, ads came to my mother again and then they left. If it weren't for the ads, they wouldn't have thought of it. (Aysun, 21)

4.2.2. Time-Saving

Time is undoubtedly a valuable resource for human beings. Therefore, consumers do not want to waste their time on commodities they find unnecessary. The sub-theme that emerged as another effective factor in the diving phase in this study is Timesaving. Participants emphasize that they want to conduct their research about any leisure service faster by using the suggested ads instead of spending time on search engines. Examining the advertisements that have been suggested based on the digital traces left by the consumer provides the consumer with a more specific review opportunity and faster access to the product with no additional research. Just as individuals do not want to waste their time, their desire to «save» time has been considered a reaction when reviewing content. One of the participant's comments for the related sub-theme is given below:

> I get furious when I open the ad and didn't find the explanations I was looking for. The cover of the ad is beautiful; the inside is a completely fiasco. I trust the cover and open the ad, stubbornly trying to reach the information about what the main page said. Then I leave finding nothing. I get frustrated for such a waste of time. (Seda, 22)

4.2.3. Leaping

When considering any event, there may be additional equipment that the consumer would not think of at first, but is actually essential for the event. While consumers are not yet aware of this situation, they can improve their experience through the equipment suggested by the ads. If additional equipment related supports the content related to it, the person is considered in the sub-theme called leaping. Leaping takes place during the diving stage, depending on the type of activity, and the participants see it as an aid. One of the participant's comments supporting this inference is: «...turns out it doesn't end with buying a tent. You need a lot of things like flashlights, stools, sleeping bags. I discovered the things I need through the ads» (Yusuf, 22).

4.3. Break-point/Inhibition

Considering the reactions toward the suggested advertising process regarding leisure services, the main factors that led to it covered the termination of the process under this theme. It also covered together user comments about the way and the kind of elements in the ads affect this process with the emotions people experience. Another factor that emerges here is that although advertisements appear to enable us to perform our purchasing behavior, the review process continues with the function of small information packages until this behavior occurs. It emphasized these phases under the diving theme. As an extension of the diving theme, the part to understand at the breakpoint is the point of terminating or skipping the ad and returning to the main feed.

4.3.1. Negative Opinion

The negative opinion topic within the break-point theme focuses on advertisements that have successfully passed the reception and driving phases but failed to satisfy the consumer and therefore could not achieve activating their purchasing behavior. Individuals stop viewing ads they find click-worthy in previous phases, because of the concepts given in this topic. The sub-topics determined for the negative opinion topics are «getting lost», «indecision», «relevance» and «locations».

Getting Lost: As a predictable extension of the comparison mentioned under the diving theme, the getting lost sub-theme manifests itself as an effort to get out because the individual gets lost in the abundance of options they encounter during the advertising process and gets bored with this situation. For the individual who is exposed to more diversity than they wish for, this diversity causes them to act negatively.

> I get exposed to so much diversity that I'm starting to quickly skip from one ad to another. The ads keep going on incessantly and I am even tired of looking, but I also want to keep doing it because I need to make plans. (Melahat, 22)

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 331-348
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

Indecision: Another reason advertising diversity causes a negative reaction is causing indecision. While continuing their research on the leisure service they are looking for, users reach different involvements by turning to different opportunities they encounter, and the diversity they end up with affects the decision-making process. The participants in the study reported that their search ends because of the indecision caused by the variety of advertisements after a certain period. The participant comment, which includes the sub-theme of indecision, is as follows:

> Let's say I'm looking for a hotel. If you have already researched once, there is no going back, it will re-appear for days. One hotel has an aqua park, the other is by the sea, and another one has another feature. Just when I decide on one of them, another suggestion appears. Then I can't decide. So, I end up leaving the ads. (Rıdvan, 21)

Relevance: The relevance of AI advertising is important to keep Ad involvement. Ads should maintain their relevance and not give the user the impression that they are ultimately reviewing them for nothing. Individuals have a negative reaction to advertisements that cannot maintain their relevance. One of the participant's comments supporting the inference is given below:

> After a certain point, absurd things happen. What I was looking for was different; the place I ended up with is different. In such cases, I get bored and turn everything off immediately. (Asuman, 21)

Locations: Although some activities can be convenient, those that take place in their city are more appealing to users. We expect although changing locations for holidays; it is often not the case for events such as concerts, festivals, or theater. In terms of holiday trips, reactions depend on whether the location meets personal expectations (such as being close to the sea, markets, clubs, etc.) «If it is close to where I am and if I have spare time, and if the price is right, I'll be there.» (Eylül, 21).

> I was in Izmir last week. I searched to see some places to go to and used my location. It's been a week since I returned, but I still get ads about different places in Izmir. There is even a place. I forgot its name. It is recommended so often that I thought that everyone in Izmir goes there. (Sude, 20)

4.3.2 Positive Opinion

AI-based advertisements, which are reacted to by the participants, together with the stimulus of the related leisure service, complete a remarkable process. When the user's reaction process is managed correctly through the themes and concepts mentioned above, they can decide to purchase leisure service products as a reflection of these reactions. «I usually go if it works for me, if I found what I was looking for, and if I still think that l need it.» (Selcuk, 20).

5. Discussion and Implications

This study investigated the reactions of consumers to AI applications in leisure services, such as travel, vacation, and entertainment. This qualitative study, which reveals the emotions, thoughts, and reactions of 23 consumers exposed to personalized AI messages or promotions in leisure services, is one of the very few studies in this line of research. Because of the conducted thematic analysis, three basic dimensions that characterize consumers' reactions to AI ads and their experiences were determined. The determined dimensions were process-based and were evaluated as reactions encountered during the processes. It entitled these themes as reception, diving, and break-point. The reception theme was evaluated in two topics as positive reception and negative reception, and the positive reception topic turned out to be related to the sub-topics of FoMo, personal interest, general knowledge, and seeking new experiences. The second theme, diving, is explained with the topics called comparison, time-saving, and leaping. The third and final theme, break-point, is divided into two topics positive and negative. The negative break-point theme was associated with getting lost, indecision, relevance, and logistics sub-topics.

While these themes that emerged because of the research were in line with some responses to artificial intelligence-based advertisements (Xian, 2021; Li, 2019), the unique side of the study was the themes or sub-themes that were not emphasized in the literature and emerged specifically in this study. For example, some observe that the reactions of the consumer to the advertisement because of FoMO and the information in the literature are parallel to each other. FoMO can be a successful strategy for starting sales and for individuals who respond positively to fear calls (Hodkinson, 2019). Indeed, Hodkinson (2019) highlighted that it started FoMO appeals through salespeople and advertisements.

Considering that it used the FoMO element in many advertisements in practice (Argan et al., 2018), some observe that this aspect of our research and the information in the literature are parallel to each other. When the responses of the participants have carefully listened to during the collection of the study data, it was observed that it related the responses given by the individuals to several stages. The first interaction with the ad includes a reaction related to the ad, and the ad responses had to be successful for them to go through the review process. Therefore, advertisements encountered on social media are clicked on or skipped depending on whether the person finds them click-worthy. It has shown that click-worthiness in social media advertisements should be evaluated in the processes, as well as exposure and involvement. With this finding, our study contributes to the subject of AI.

The principal managerial contribution of this study is to show that the crucial point in a succession of AI ads is to contain creative content which could attract the attention of the consumer. The results highlight that the manager needs to consider should evaluate both positive and negative aspects of AI from a consumer-oriented perspective. It should have considered that it is necessary to be careful in matters that may disturb the consumer and are likely to intrude into private areas' aspects of their lives. For AI advertising massages not to go to waste, marketing managers should consider the issue from the perspective of consumer behavior. It related selectivity in perception to a desire to click and see AI advertising.

As shown by 'psychological variables in consumer behavior', perception increases when the message exactly matches the person's needs. Within the framework of consumer behavior research, this study reveals how variables, such as perception and learning, which are considered intrinsic or psychological factors, affect AI advertising. Understanding the users' reactions to AI ads gives managers another perspective based on which they can improve the ad creation process by creating more relevant and engaging ad messages that improve the customer experience.

This study also shows that AI algorithms can predict consumer expectations and desires on a large scale and that they can apply consumer behavior theories and variables to improve advertiser-user interactions. Thus, using theories such as personalization, needs, expectations, and motivation to attract the attention of consumers provides managers with opportunities.

6. Conclusion

This study examines the behavior of social media users exposed to AI ads. Using a qualitative study, which included an in-depth interview, we found three process themes that influence consumer reactions to AI-based advertisement messages among social media users: reception, diving in, and break-point. The first stage comprises positive and negative sub-themes. If the AI ads attract the attention of a consumer, positive evaluation occurs and the consumer's evaluation of AI ads continues. That this positive stage, FoMO, personal interest, general culture, and seeking experience come into play. The second stage, entitled diving in, includes sub-themes, comparison, time-saving, and leaping. The final stage represents the decision-making stage and includes negative or positive opinions. Regarding theoretical implications, AI-based brand ads need to constantly update their knowledge of the psychological mechanisms that will motivate consumers. The current study contributes to social media user reactions to artificial intelligence-based advertisements. Specifically, our study contributes to the literature on the success criteria of AI-based advertising by demonstrating that psychological and impulsive factors have significant effects on the reactions of users using social media and being exposed to AI ads. Considering managerial implications, brands that use AI ad technologies should put themselves in the place of the consumers. In other words, they should empathize. A synthesis of consumer reactions on social media is important in advertising, as well as AI research and practice because a user-based perspective is required to achieve effective results.

7. Limitation and Future Studies

As in every scientific research, our study also has some limitations. We can consider this study pioneering research, as it provides the first qualitative, in-depth descriptions of consumers' reactions to AI ads. To eliminate the limitations of research, the validity and reliability issues in qualitative research, such as member checks, are emphasized. However, our research still has some limitations, as in every qualitative research. It may limit the generalizability of these results because of the small sample size. Therefore, the results of our study should be evaluated in this limitation. To get more generalizable results, it would be beneficial to include participants from more countries and perform quantitative studies. It can achieve a broader and generalizable perspective by using consumer-based scales related to AI advertising metrics or by developing new scales.

8. Acknowledgements

The authors thank Dr. Adam Stone for proofreading voluntarily this manuscript.

References

Aaker, D. A., and Stayman, D. M., 1992. Implementing the concept of transformational advertising. *Psychology & Marketing*, 9(3): 237-253.

Argan, M., Argan, M. T., and İpek, G., 2018. I Wish I Were! Anatomy of a fomsumer. *Journal of Internet Applications and Management*, 9(1): 43-57.

Ashraf, S., Gao, M., Chen, Z., Naeem, H., Ahmad, A., and Ahmed, T., 2020a. Underwater pragmatic routing approach through packet reverberation mechanism. IEEE Access, 8, 163091-163114.

Ashraf, S., Muhammad, D., Shuaeeb, M., and Aslam, Z., 2020b. Development of Shrewd Cosmetology Model through Fuzzy Logic. *Journal of Research in Engineering and Applied Sciences*, 5(3): 93-99.

Batra, R., and Keller, K. L., 2016. Integrating marketing communications: New findings, new lessons, and new ideas. *Journal of Marketing*, 80(6): 122-145.

Braun, V., and Clarke, V., 2006. «Using thematic analysis in psychology». *Qualitative research in Psychology*, 3(2): 77-101.

Campbell, C., Plangger, K., Sands, S., and Kietzmann, J., 2022. Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. *Journal of Advertising*, 51(1): 22-38.

Charmaz, K., 2006. *Constructing grounded theory, a practical guide through qualitative analysis*. London: Sage.

Choi, J. A., and Lim, K., 2020. Identifying machine learning techniques for classification of target advertising. *ICT Express*, 6(3): 175-180.

Churcher, P. R., 1991. The impact of artificial intelligence on leisure. *AI & Society*, 5(2): 147-155.

Creswell, J. W., 2013. *Qualitative enquiry & research design, choosing among five approaches*. 3rd ed. Los Angeles, CA: Sage.

Glaser, B. G., and Strauss, A. L., 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Hodkinson, C., 2019. 'Fear of Missing Out'(FOMO) marketing appeals: A conceptual model. *Journal of Marketing Communications*, 25(1): 65-88.

Hou, C. I., 2013. Traffic flow forecasting in leisure farm areas using artificial neural networks. *Przegląd Elektrotechniczny*, 89(1b): 205-207.

JWT Intelligence., 2012. Fear of missing out (FOMO). Retrieved November 28, 2017 from JWT: http://www.jwtintelligence.com/wpcontent/uploads/2012/03/F_JWT_FOMO-update_3.21.12.pdf

Kietzmann, J., and Canhoto, A., 2013. Bittersweet! Understanding and managing electronic word of mouth. *Journal of Public Affairs*, 13(2): 146-159.

Kietzmann, J., Paschen, J., and Treen, E., 2018. Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research*, 58(3): 263-267.

Lai, Z., 2020. Research on Advertising Core Business Reformation Driven by Artificial Intelligence. *Journal of Physics: Conference Series:* 1-9.

Lashua, B. D., 2014. DWYL? YOLO. *Annals of Leisure Research*, 17(2): 121–126.

Lashua, B. D., 2018. The time machine: Leisure science (Fiction) and futurology. *Leisure Sciences*, 40(1-2): 85-94.

Li, H., 2019. Special section introduction: Artificial intelligence and advertising. *Journal of advertising,* 48(4): 333-337.

Li, H., Edwards, S. M., and Lee, J. H., 2002. Measuring the intrusiveness of advertisements: Scale development and validation. *Journal of Advertising*, 31(2): 37-47.

Malterud, K., Siersma, V. D., and Guassora, A. D., 2016. Sample size in qualitative interview studies: guided by information power. *Qualitative Health Research*, 26(13): 1753-1760.

McCarthy, J., and Hayes, P. J., 1981. Some philosophical problems from the standpoint of artificial intelligence. In *Readings in artificial intelligence* (pp. 431-450). Morgan Kaufmann.

Mogaji, E., 2018. *Emotional Appeals in Advertising Banking Services*. London: Emeald.

Murgai, A., 2018. Transforming digital marketing with artificial intelligence. *International Journal of Latest Technology in Engineering, Management & Applied Science*, 7(4): 259-262.

Przybylski, A. K., Murayama, K., DeHaan, C. R., and Gladwell, V., 2013. Motivational, Emotional, and Behavioral Correlates of Fear of Missing Out. *Computers in Human Behavior*, 29(4): 1841-1848.

Rodgers, S., 2021. Themed issue introduction: Promises and perils of artificial intelligence and advertising, *Journal of Advertising*, 50(1): 1–10.

Rodgers, W., and Nguyen, T. (2022). Advertising benefits from ethical artificial intelligence algorithmic purchase decision pathways. *Journal of Business Ethics*: 1-19.

Shah, N., Engineer, S., Bhagat, N., Chauhan, H., and Shah, M., 2020. Research trends on the usage of machine learning and artificial intelligence in advertising. *Augmented Human Research*, 5(1): 1-15.

Shumanov, M., Cooper, H., and Ewing, M., 2021. Using AI predicted personality to enhance advertising effectiveness. *European Journal of Marketing*: 1-20.

Vakratsas, D., and Wang, X., 2020. Artificial intelligence in advertising creativity. *Journal of Advertising*, 50(1): 39-51.

Wu, L., Dodoo, N. A., Wen, T. J., and Ke, L., 2021. Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing. *International Journal of Advertising*: 1-18.

Xian, X., 2021. Psychological Factors in Consumer Acceptance of Artificial Intelligence in Leisure Economy: A Structural Equation Model. *Journal of Internet Technology*, 22(3): 697-705.

Zhao, H., Lyu, F, and Luo, Y., 2020. Research on the effect of online marketing based on multimodal fusion and artificial intelligence in the context of big data*.* *Security and Communication Networks:* 1-9.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 349-365
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.28435

Regression Based Performance Analysis and Fault Detection in Induction Motors by Using Deep Learning Technique

Pradeep Katta[a], Karunanithi Kandasamy[b], Raja Soosaimarian Peter Raj[c], Ramesh Subramanian[d], and Chandrasekar Perumal[e]

[a] Research Scholar, School of Electrical and Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

[b,d,e] Professor, School of Electrical and Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

[c] Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India

✉ pradeep.2048@gmail.com, k.karunanithiklu@gmail.com, avemariaraja@gmail.com, rameshsme@gmail.com, drchandrasekar@veltech.edu.in

KEYWORDS

induction motor; DBN; RBM; Fast Fourier Transform (FFT); regression modeling

ABSTRACT

The recent improvements related to the area of electric locomotive, power electronics, assembly processes and manufacturing of machines have increased the robustness and reliability of induction motors. Regardless of the increased availability, the application of induction motors in many fields alleges the need for operating state supervision and condition monitoring. In other words, fault identification at the initial stage helps make appropriate control decisions, influencing product quality as well as providing safety. Inspired by these demands, this work proposes a regression based modeling for the analysis of performance in induction motors. In this approach, the feature extraction process is combined with classification for efficient fault detection. Deep Belief Network (DBN) stacked with multiple Restricted Boltzmann Machine (RBM) is exploited for the robust diagnosis of faults with the adoption of training process. The influences of harmonics over induction motors are identified and the losses are mitigated. The simulation of the suggested approach and its comparison with traditional approaches are executed. An overall accuracy of 99.5% is obtained which in turn proves the efficiency of DBN in detecting faults.

1. Introduction

Induction motors (IM) are regarded as the major load in electric power systems and contributes to half of the electric energy demand consumed globally (Donolo et al., 2020). These motors are incorporated equipment which are generally utilized in industrial applications and manufacturing processes. The proper functioning of IMs is mandatory for the effective functioning of industries. In contrast, numerous faults occur frequently in Ims because of enduring or overrated loads, tough working conditions, regular wear and tear as well as unexpected events. This requires an efficient fault detection approach to prevent shutdown, unsafe operation, repair as well as operational costs and catastrophic failures (Liu et al., 2017). Moreover, IMs are subjected to numerous power quality problems like imbalance, fluctuations and disturbances in voltage. Additional losses occur when IMs are supplied with voltage of decreased quality. This further results in decreased efficiency and overheating (Ghaseminezhad et al., 2017; Amaize et al., 2017).

Generally, under conditions where voltage is unbalanced, the equivalent induction motor circuit is separated into negative and positive sequence equivalent circuits. The distorted waveform of voltage obtained during this condition has large harmonic components which further performs distortion of current and flux density waveforms in the induction motor. This results in the efficiency being compromised and in complicated loss characteristics during distorted as well as unbalanced voltage conditions. The analysis of induction motor loss considers namely, the rotor electromagnetic quantity harmonics have to initially identified. At partial-load condition, in the induction motor, the fundamental frequency of the rotor is less than 1 Hz. This signal of reduced frequency comprises of numerous power supply cycles incurring greater calculation cost. Hence, rapid analysis of induction motor loss characteristics is tedious. Secondly, iron loss prediction is complicated in an induction motor due to the magnetic material characteristics which are nonlinear in nature. A finite element model (FEM) which exhibits extensive improvements, is applied to an induction motor to determine loss. Unfortunately, the theoretical study related to distortion and voltage imbalance on induction motor loss properties is not sufficient (Zhang et al., 2018). Hence, regression based analysis is exploited in this work which solves the power quality issues and mitigates the induction motor losses.

In induction motors, fault detection is important to maintain the efficient functioning of industrial process without interruption (Ali et al., 2019). The three basic approaches for fault detection in induction motors are mentioned as follows: i) the signature extraction based approach (Hassan et al., 2018) is performed by the survey of fault signatures in frequency or time domain. This approach requires more advanced tools for signal processing. ii) The model based approach (Su et al., 2017) depends on mathematical modeling for predicting induction motor behaviour during fault conditions. Even though this approach generates warnings as well as determines incipient faults, the accuracy depends largely on explicit motor types which are scarcely available. iii) The knowledge based approach (Xu et al., 2017) does not require characteristics of motor or load, trigger threshold and machine models. On the contrary, this approach demands thorough data analysis to generate effective rules or inferences for fault detection. The Discrete Wavelet Transform (DWT) approach (Bessous et al., 2018),(Ghods et al., 2016; Nishat Toma et al., 2020) performs efficiently with increased power motors during steady-state circumstances including constant torque, speed and voltage. It identifies the harmonics thereby distinguishing the induction motor faults during steady state. In contrast, with non-stationary signals, the deterioration of spectrum occurs, resulting in information loss of fault diagnosis. Moreover, the DWT based approach depends heavily on the quality of the measurement, the number of samples taken as well as sampling frequency. Henceforth, to improvise the monitoring and diagnostic performance of an induction motor, machine learning approaches are opted. Few of them are Random forest (Ma et al., 2018), Support Vector Machine (Pandarakone et al., 2016), K-nearest neighbour (Dong et al., 2016), Decision Tree (Toma et al., 2020) and Naive Bayes classifier (Zhang et al., 2018).

The adoption of these Machine Learning (ML) approaches results in decreased reliability due to more missing data. The introduction of deep learning approaches addresses these shortcomings and generates acceptable and promising results related to fault diagnosis in induction motors. These approaches utilized manually extracted features (Abid et al., 2019) . Some of the merits of deep learning approaches over ML approaches are as follows: versatility to be trained for any type of utility, reduced computation time as well as parameter tuning in the absence of programming skill. Among the deep learning approaches, Convolutional Neural Networks (CNN) is utilized for the diagnosis of faults but did not provide the deep analysis (Ding et al., 2017; Hsueh et al., 2019).

Considering the above discussed factors, this work contributes an approach based on deep learning for the mitigation of losses and fault detection in induction motors. A regression based analysis is performed for estimating the power factor at varying load because reduced power factor generates huge losses in induction motor. The DBN comprising of RBM is adopted for the attainment of fault detection in induction motors with improved classification accuracy. The remaining paper is structured as follows: Section overviews related works, Section 3 presents the proposed approach, Section 4 outlines the results and a discussion, Section 5 draws conclusions from the conducted research.

2. Related Work

Prashant Kumar and Ananda Shankar Hati (Kumar et al., 2021) have introduced an adaptive gradient optimizer based deep convolutional neural network (ADG-dCNN) approach for finding the faults occurring in the rotor and bearing of squirrel cage induction motor. By this approach, human expertise requirement is highly reduced and the feature extraction process is enabled automatically. For fault detection, vibration data from numerous MEMS accelerometers are gathered. Here, SHapley Additive exPlanations (SHAP) technique has been implemented for evaluating fault classification. It is difficult to carry out multiple fault detection since a large amount of data is required in order to achieve high fault categorization accuracy.

R. R. Kumar et al (2021) have proposed a new technique for identifying and categorising faults in an induction motor with the help of a two stage Neural Network (NN). Here all data are represented by using Principle Component Analysis (PCA) which is a linear-based feature reduction method. The three- phase current signals are detected and classified online using shallow neural networks. Overfitting inhibits the generalization capacity.

S. M. K. Zaman et al (2021) have introduced a Graph-based Semi Supervised Learning (GSSL) method for identifying the faults that occur in a direct online induction motor. The prominent advantage of GSSL is that the training datasets require a lower number of labelled data. The Greedy-Gradient Max Cut (GGMC) algorithm from the GSSL family has been used in this research to analyse the compatibility of this approach. Obtaining a huge labelled training dataset is quite costly, difficult and time consuming.

A. Chouhan et al (2021) have developed an effective Artificial Neural Network (ANN) model for predicting both electrical and mechanical faults in an induction motor, which is operated at three different conditions: same speed, interpolated speed and extrapolated speed. A machinery fault stimulator is utilized to measure the current and vibration signals of healthy and faulty induction motors. This fault detection technique is more accurate in interpolation speed cases than extrapolation speed cases.

A. Ibrahim et al (2022) have proposed a novel Hybrid Invasive Weed Optimization (IWO) method for the detection of induction-motor faults. An optimal features selection process was implemented and machine learning classifiers were trained to classify the faults. Here, fault identification had not been carried out at the early stage by taking into account the current and vibration signal.

3. Proposed Approach

The performance analysis as well as fault detection in induction motors have evolved as a significant area of interest. Numerous researches have been carried out in relation to the monitoring of induction motor operation and the early detection of faults. From among them, deep learning based approaches are regarded as efficient and in this paper, DBN comprising of multiple stacks of RBM has been utilized. The block diagram of the suggested approach is given in Figure 1.

The data from the induction motor at varying conditions of rotating speed has been obtained and subjected to regression analysis. It is utilized to estimate the appropriate power factor in order to mitigate the losses occurring in induction motors. Further, DBN has been adopted for the identification of faults with an improved training of RBM. In general, DBN comprises of multiple hidden layers with the ability to learn hierarchical representation in an unsupervised manner. It also performs classification and RBM is exploited to make the learning of abundant parameters easier.

3.1. Regression Modeling

Regression is regarded as a statistical approach utilized for the fitting of nonlinear as well as linear models. It denotes the relation between the independent variable p and the dependent variable q considering a n^{th} degree polynomial. Regression involves two main factors, given as: Consideration of observed values to simulate a regression model and prediction of polynomial coefficients. The corresponding expression is given by,

$$q_i = \sum_{i=1}^{n} f(p_i, \beta) + \varepsilon_i \tag{1}$$

$$\hat{q}_i = f(p_i, \beta) = \beta_0 + \beta_1 p_1 + \beta_2 p_2^2 + \quad + \beta_m p_n^m \tag{2}$$

Here, q_i indicates observed dependent variable, $f(p_i, \beta)$ indicates predicted value, ε_i represents the error between the estimated and observed value. $(\beta_0, \beta_1, \beta_2, ..., \beta_m)$ are the polynomial coefficients in which m indicates the coefficient number. $p_1 + p_2^2 + ... p_n^m$ are independent variables in which m

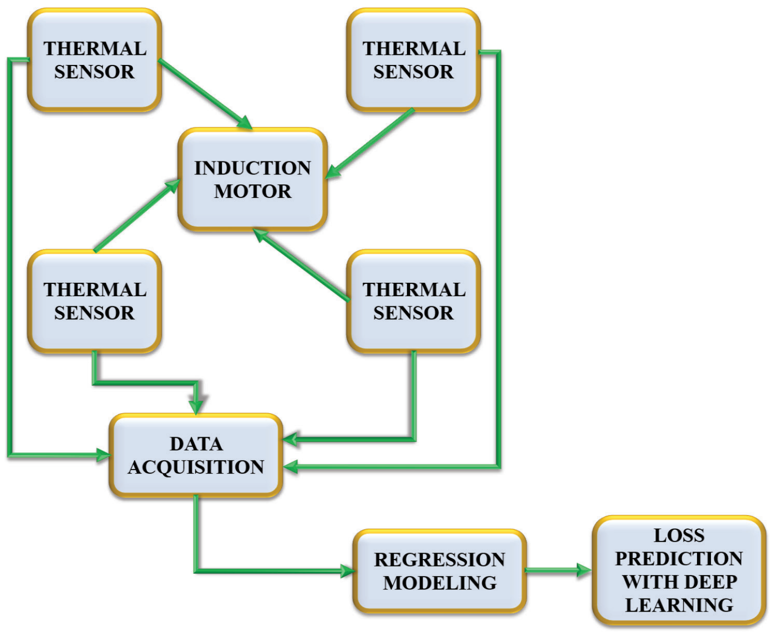


Figure 1. Process Flow Diagram

represents the polynomial degree number and n represents the number of variables. The difference between (1) and (2) provides the value of the mean square error. The sensors provided in the system perform observation of speed as well as motor torque. The power factor is obtained from the equation given below.

$$Load\ \ factor = \frac{PF \times \sqrt{3} \times V \times I}{P_{in}} \times 100\% \tag{3}$$

$$P_{in} = \frac{0.7475hp}{\eta} \tag{4}$$

Here, V indicates the rms voltage, I is the rms current, PF represents the power factor, η denotes efficiency and P_{in} indicates input power at full rated load. Since the power factor of the induction motor is nonlinear, the regression model has been adopted to solve and determine the power factor value.

Figure 2 denotes the parameters measured including the power factor, total current, active current and reactive current. The load factor is estimated by equations (3) and (4) from which the power factor is calculated. It is clear that the power factor is low at minimized load condition because during this condition, the motor consumes only reactive current and the active current approximately equals zero. The active current increases with the increase in load which further results in the increase of the power factor. Hence, in conditions where load is reduced, the power factor has to be detected because low power factor results in losses. The regression based modeling is regarded as a solution of reduced cost for the determination of the power factor at varying load condition.

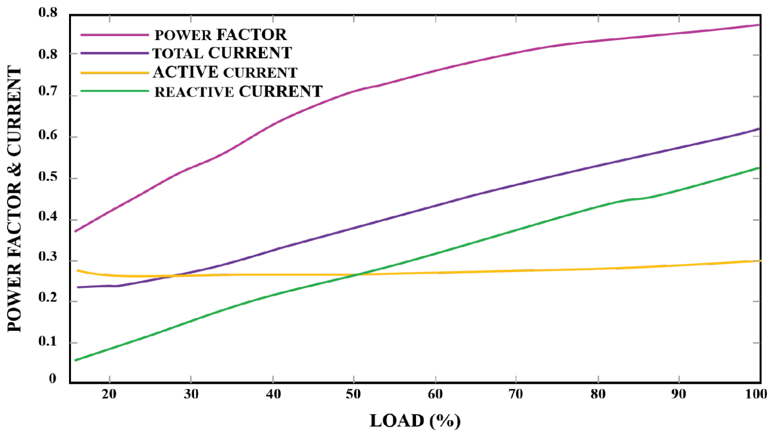


Figure 2. Measurement of the current and power factor

3.2. RBM Architecture and Training

The RBM is utilized as generative model of various data types including signal coefficients and performs learning approach for DBN. Usually, RBMs are trained utilizing the learning procedures which considers factors such as momentum, learning rate, number of hidden units, mini-batch size and initial value of weights. The architecture of RBM given in Figure 3 comprises a visible layer and a hidden layer. All the visible layer units are connected to the hidden layer units but the units are not interconnected. Considering the hidden and visible units, the energy of (v, h) is given by,

$$E(v,h) = -\sum_{i\in visible} x_i v_i - \sum_{j\in hidden} y_i h_i - \sum_{i,j} v_i h_j w_{ij} \tag{5}$$

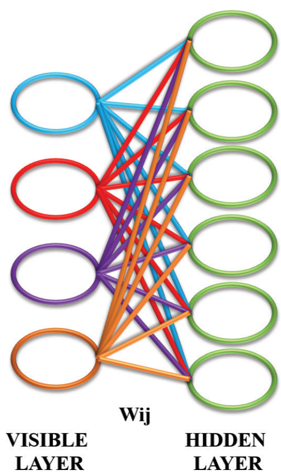


Figure 3. RBM architecture

Here, v_i, h_j denote visible unit as well as hidden unit binary states, x_i and y_i represent the biases, w_{ij} denote the weight between the binary states. A probability is assigned to each possible pair of hidden and visible layers through the function given by,

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \tag{6}$$

The partition function Z is given by,

$$Z = \sum_{v,h} e^{-E(v,h)} \tag{7}$$

Probability of assigning a network to the visible vector v is denoted by the summation of all hidden vectors which are possible.

$$p(v) = \frac{1}{Z} \sum_{h} e^{-E(v,h)} \tag{8}$$

The probability of assigning a network to a training data is increased with the adjustment of weights as well as biases for reducing energy of that specific data. This is followed by increasing the energy of other data, specifically data with reduced energy. The log probability derivative of training vector with respect to weight is given by,

$$\frac{\partial \log p(v)}{\partial w_{ij}} = \left[v_i h_j\right]_{data} - \left[v_i h_j\right]_{model} \tag{9}$$

The simple learning rule for the log probability of training data is given by,

$$\Delta w_{ij} = \varepsilon([v_i h_j]_{data} - [v_i h_j]_{model}) \tag{10}$$

Here ε denotes the learning rate. The unbiased sample of $v_i h_{jdata}$ is obtained easily due to the absence of direct connection within hidden units of RBM. When a training data v is selected randomly, the binary state h_j of every hidden unit is set to 1 with probability given by,

$$p(h_j = 1 \mid v) = \sigma\left(y_j + \sum_i v_i w_{ij}\right) \tag{11}$$

Here, $\sigma(a)$ indicates a logistic function $1/(1 + exp\,(-x))$

Similarly, the unbiased sample of visible units is easily obtained due to the absence of direct connection between visible units.

$$p(v_i = 1 \mid h) = \sigma\left(x_j + \sum_j h_j w_{ij}\right) \tag{12}$$

To obtain an unbiased sample of *vihjmodel*, any random state of visible unit is initialized and Gibbs sampling is performed alternatively. A single iteration comprises of updating hidden unit using equation (11) subsequently followed by updating visible unit using equation (12).

3.3. DBN Architecture

DBN is a deep learning network comprising of numerous hidden layers with many nonlinear representations. This network is regarded as a probabilistic generative model in which multiple RBMs are stacked together as shown in Figure 4.

Since, multiple hidden layers are present in DBN, the learning is carried out from input data and hierarchical representation related to every hidden layer is extracted. The joint distribution within l hidden layers h_k and visible layer v is estimated as,

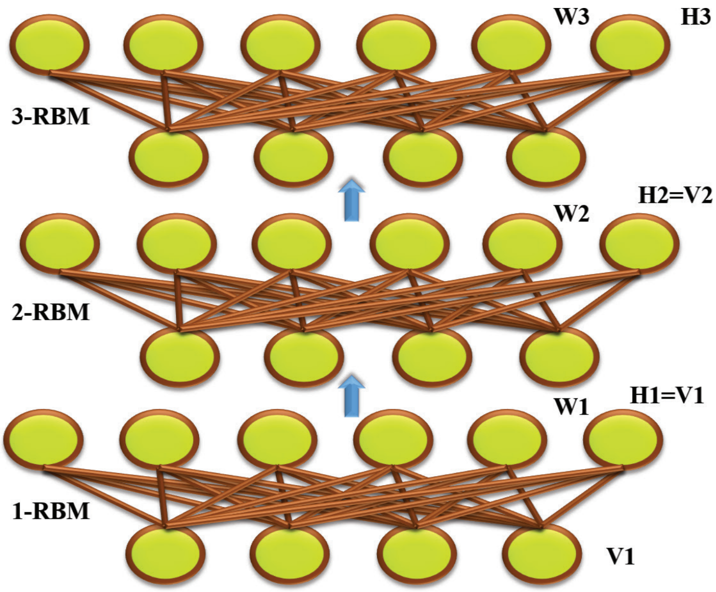


Figure 4. DBN architecture

$$P(v,h^1,\dots\quad,h^m)=\left(\prod_{k=1}^{m-1}P\left(h^{k-1}\mid h^k\right)\right)P\left(h^{m-1},h^m\right) \tag{13}$$

The method of learning a wide range of parameters utilizing conventional learning methods is tedious due to the transmission of errors to lower levels. Added to this, the ability for parameter adjustment is complicated resulting in difficulties for the generation of optimal parameters. Due to this, the following procedure is adopted for the training purpose: Initially, the hidden layer and input units are trained by the RBM rule. Next, the first hidden layer output is given to the second hidden layer input. Similarly all hidden layers are trained till the training of all hidden layers is completed.

3.4. DBN for Fault Diagnosis

A fault detection approach based on DBN comprising of RBM is proposed for induction motors as shown in Figure 5. DBN has been chosen due to its supervised learning and its ability to reduce training error with through improved classification accuracy.

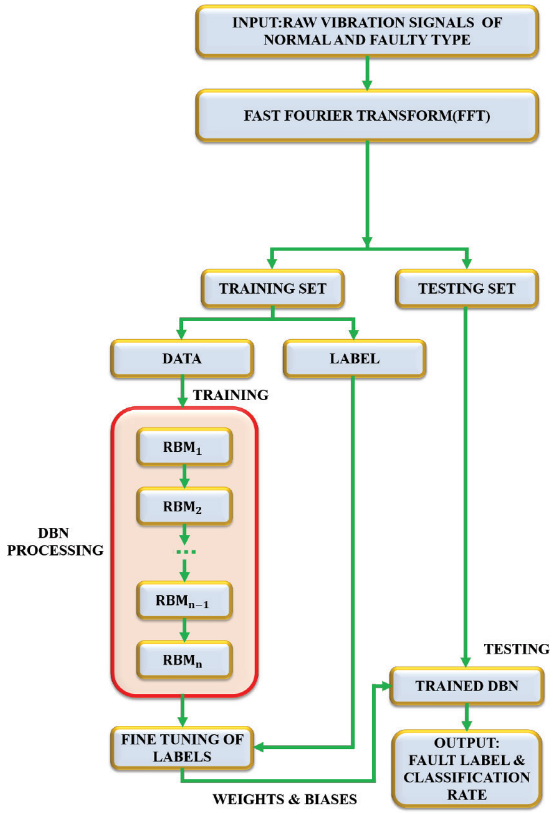


Figure 5. Fault detection approach based on DBN

Since the vibration signals contain significant information reflecting the working condition of induction motors, these signals are chosen as input signals. Due to the presence of correlation among sampled data points, the adjacent classification tasks get affected. To avoid this, the vibration signals are converted into frequency domain from time domain utilizing FFT. This conversion enables the evaluation of difference within constitutive frequency components and their distribution. In this work, an amplitude of $30Hz$ and $90Hz$ components is selected as features for the categorization of faults in induction motors. This is because, at the above mentioned frequencies, high difference in amplitude is obtained which differentiates the operating conditions of the induction motor. The feature location generally depends on the rotation speed of the induction motor. During analysis, no overlapping occurs and hence the faults are identified at the initial stage. Each of the obtained features is applied as input to the DBN architecture and the corresponding training process is explained below.

Step 1: Initialize the input parameters and input data set.

Step 2: Initialize the first layer for training such that $i = 1$.

Step 3: Train the first layer using the RBM learning rule.

Step 4: Save the obtained representations, biases and weights.

Step 5: Perform training till the completion of all layers.

Step 6: Following training, perform supervised learning for the classification of faults.

Step 7: Perform fine tuning of all parameters.

Thus, the DBN is exploited for efficient training of input data thereby providing accurate classified results in the detection of faults in induction motors. The deep network parameters are updated on the bases of learning rules for the generation of more accurate results.

4. Results and Discussion

The experimentation and simulation of the proposed approach considers an induction motor. The behavior of the motor related to different faults has been analysed utilizing the fast fourier transform. It converts the variations of stator current into domain and these features are treated for the deep learning approach for fault diagnosis and classification. The FFT feature dataset comprises three partitions: i) A set of 2400 data for training, fine tuning and testing. ii) A set of 800 data without fault types for training which has been utilized to perform forward stack-learning of RBM from input to high layers. iii) A set of 800 data with fault types has been used to investigate DBN's recognition rate. It performs learning in backward fine-tuning manner from the classification layer to the low layers.

Figure 6 represents the waveform indicating the speed of the motor. The induction motor was made to operate in faulty conditions and the corresponding motor speed was retrieved. Subsequently, the power factor is estimated to determine losses and this is performed by regression modeling. In this approach, polynomial function is applied to obtain the power factor value with high accuracy.

The coefficients of the polynomial are predicted and the unknown power factor values are determined. This approach provides improved performance with increased flexibility and reliability. The accuracy of the unknown estimation points is determined by the residual errors of set points. Figures 7 (a), 7 (b) and 7(c) represents the estimated power factor curve and the measured power factor curve with residual errors of 0.03 for first order, 0.001 for second order, 0.000130 for third order. The obtained outputs clearly indicate that the residual error decreases with the increase in polynomial order.

The utilization of actual signals in the absence of pre-processing does not provide accurate classification. In order to achieve accurate classified results, appropriate features are extracted for fault detection. The outputs rely on the extracted features' quality and hence in this approach, feature learning as well as classification are combined together by DBN to improve the efficiency of fault detection in induction motors. FFT is exploited for the extraction of appropriate features which are further subjected to classification. The training of DBN is carried out utilizing mean square error as the measurement of performance calculation. An improved performance of 0.00262 is obtained with the consideration of 200 epochs as shown in Figure 8.

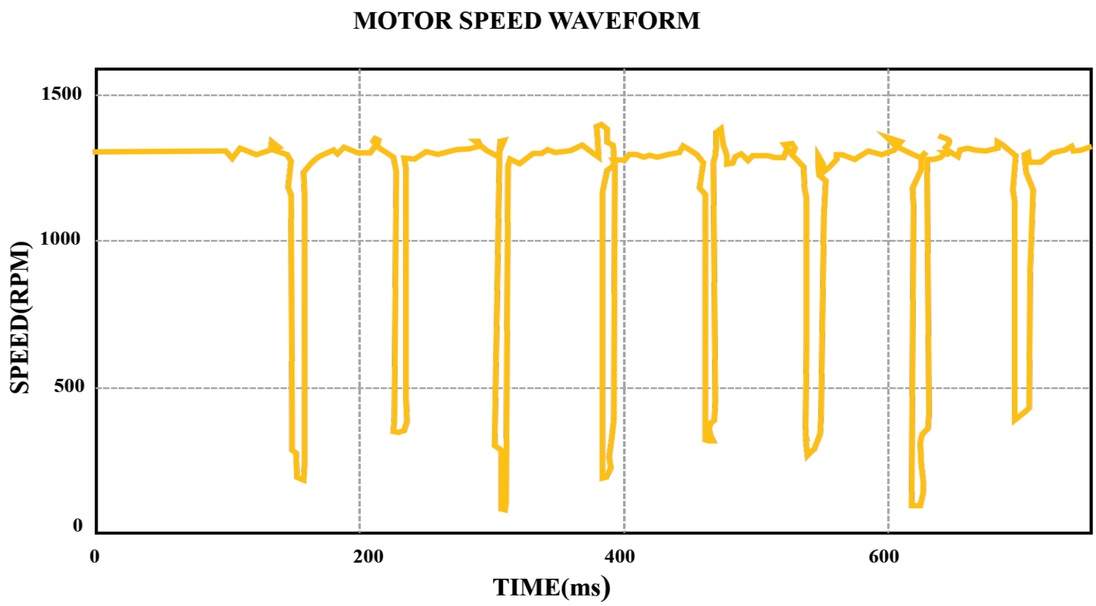


Figure 6. Waveform for motor speed

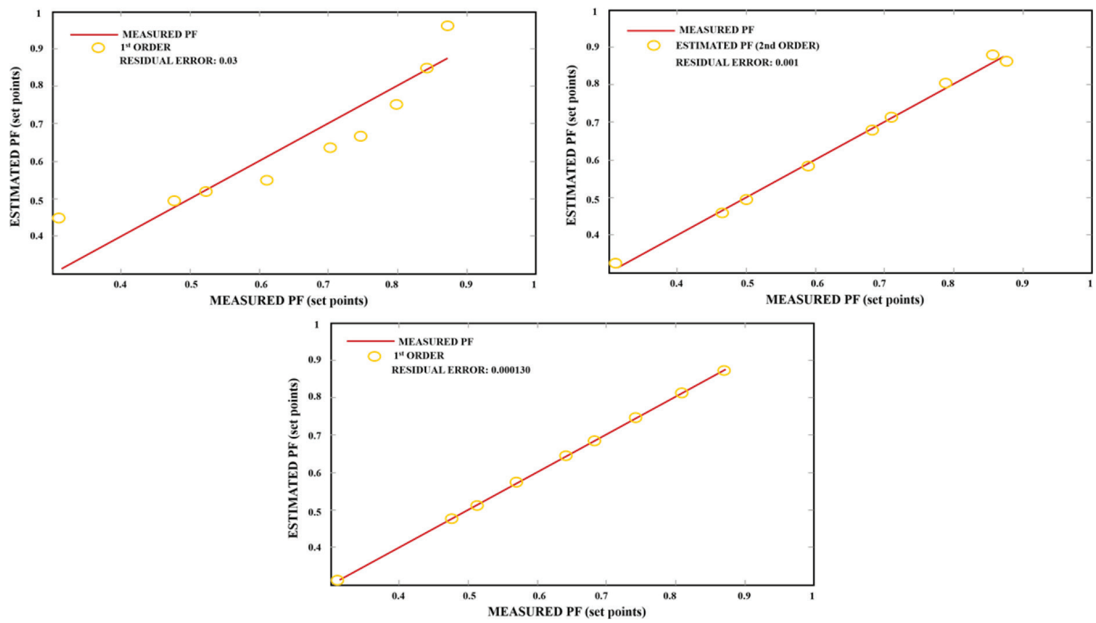


Figure 7. Residual error of (a) first order (b) second order (c) third order polynomial

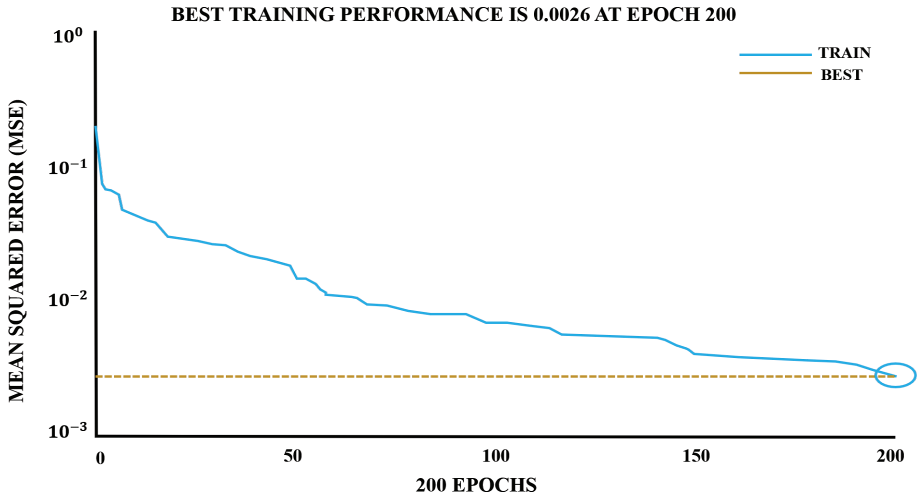


Figure 8. Training of DBN

Figure 9 indicates the Receiver Operation Characteristic (ROC) curve of validation and testing. The obtained lines for each fault reaches 1 and this indicates reduced false positive rate with increased system accuracy. Figure 10 (a) indicates the accuracy and 10 (b) indicates the loss of the proposed system. The rate of accuracy increases with the epoch and attains an increased value of 99.5%. The choice of epoch is regarded as a crucial factor because it generates a critical point at which overfitting of network to training data occurs.

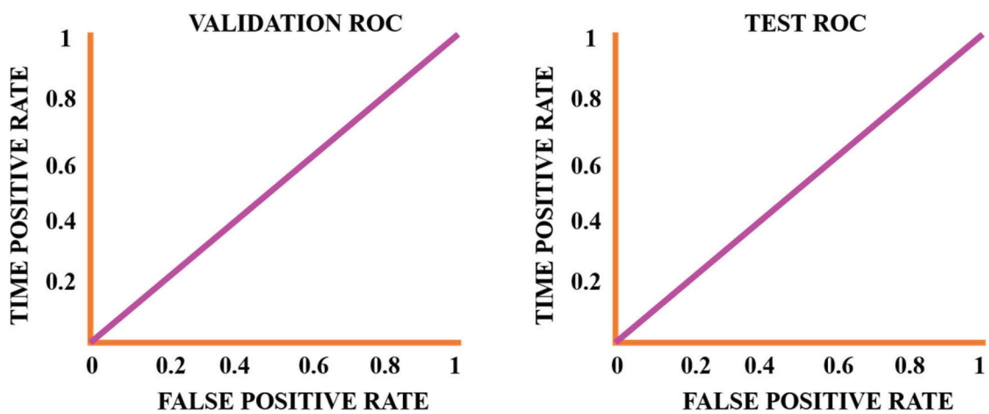


Figure 9. ROC curves of validation and testing

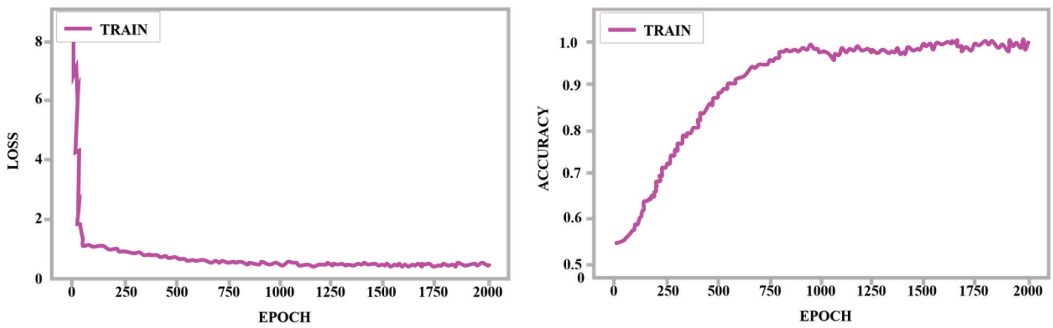


Figure 10. DBN training curve (a) Loss (b) Accuracy

Table 1 indicates the classification report of the test dataset which clearly indicates that the proposed DBN performs efficacious extraction and learning of features from the dataset. It performs efficient classification of healthy state, stator and rotor friction, bearing axis deviation, poor insulation, bearing noise and rotor aluminium end ring break. The performance of the proposed fault identification model has been analysed on the basis of sensitivity, precision, recall and F1 score.Moreover, the assessment of the aforementioned factors helps to find the industrial applicability of the proposed model.

Table 1. Test dataset classification report

Class	FI Score	Precision	Recall	Sensitivity
Healthy	1.00	1.00	1.00	1.00
Stator and rotor friction	0.96	0.95	0.97	0.95
Bearing axis deviation	0.94	0.97	0.96	0.94
Poor insulation	0.95	0.96	0.89	0.90
Bearing noise	0.96	0.93	1.00	0.95
Rotor aluminium end ring break	1.00	1.00	1.00	1.00

Table 2 represents a comparison of performance measures such as mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE) and normalized absolute error (NAE) of Support Vector Regression (SVR) and Back propagation (BP) with DBN. The proposed DBN based approach generated reduced values of error and is highly advantageous when compared to other existing approaches.

Table 2. Performance measure comparison

Approach	MAPE	RMSE	MAE	NAE
SVR	1.032	0.016	0.010	0.015
BP	1.583	0.028	0.016	0.018
DBN	0.928	0.015	0.009	0.010

Figure 11 represents the comparison of training error of the proposed DBN with other existing approaches such as SVR and BP. The proposed DBN is found to be more efficient with reduced training error when compared to other approaches.

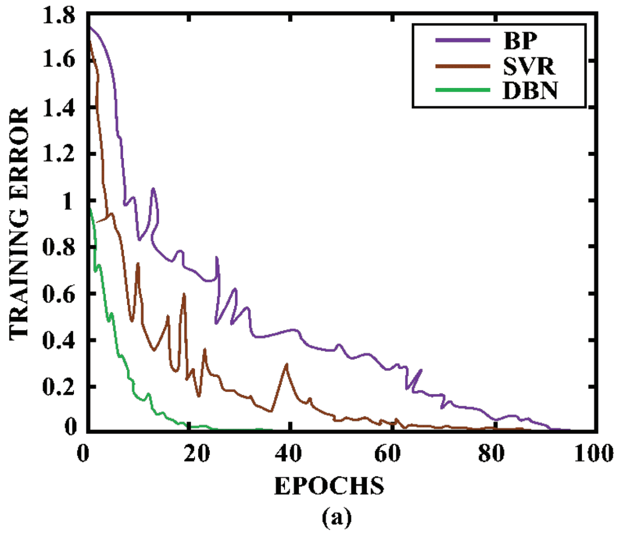


Figure 11. Comparison of training error

Figure 12 indicates the comparison of overall accuracy of the proposed DBN with other existing classifiers such as K-nearest neighbour, Multi-layer perceptron and support vector machine. The comparative reveals that the proposed approach generated improved classification results when compared to other approaches with an accuracy of 99.5%.

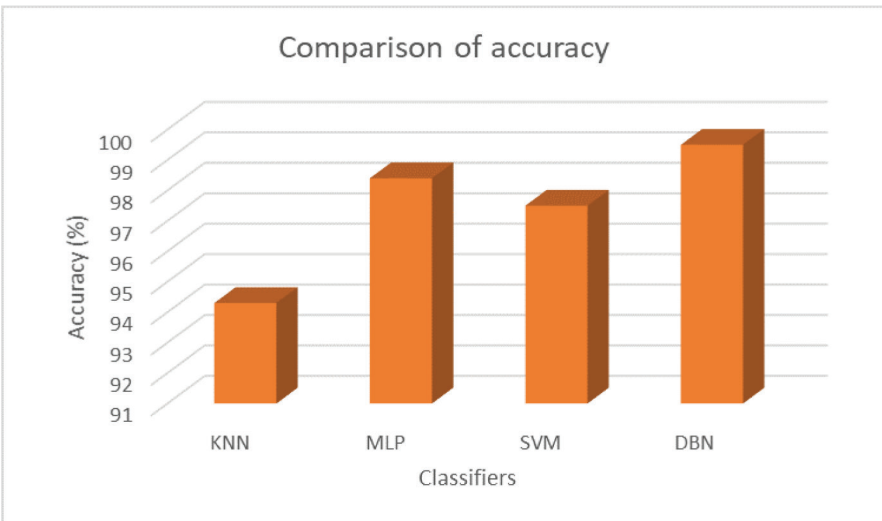


Figure 12. Comparison of accuracy

5. Conclusions

A novel approach based on deep learning has benn proposed for the mitigation of losses and fault diagnosis in induction motors. It applies regression based analysis and adopts DBN for the efficient detection of faults. DBN is formed by the stacking of RBM units one over the other. FFT is utilized for the conversion of input time domain signals to frequency domain which is further trained by DBN. The proposed approach minimizes the training error as well as enhances the accuracy of classification with the learning of multiple representation layers.The Obtained simulation results show that the deep learning based DBN is efficient in the detection of induction motor faults and generates an accuracy of 99.5%. Thus, the proposed approach generates improved classification results when compared to other traditional approaches. Future work related to the proposed approach may concentrate on the issue of overfitting. This in turn can lead to greater improvements in the performance of the DBN model in the diagnosis of faults.

5.1. Conflict of interest

The authors have no conflict of interest in publishing this work.

5.2. Plagiarism

The work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), and is not under consideration for publication elsewhere. No article with the same content will be submitted for publication elsewhere while it is under review by ADCAIJ, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, without the written consent of the copyright-holder.

5.3. Contributors

All authors have materially participated in the research and article preparation. All authors have approved the final article to be true.

References

Abid Ben, F., Sallem, M., and Braham, A., 2019. Robust interpretable deep learning for intelligent fault diagnosis of induction motors, IEEE Transactions on Instrumentation and Measurement, Vol: 69, No: 6, pp: 3506-3515.

Ali Zawad, M., Shabbir, M. N. S. K., Liang, X., Zhang, Y., and Hu T., 2019. Machine learning-based fault diagnosis for single-and multi-faults in induction motors using measured stator currents and vibration signals. IEEE Transactions on Industry Applications, Vol: 55, No: 3, pp: 2378-2391.

Amaize, P., Ignatius, K. O., Oluwasogo, E. S., Alayande, A. S., and Airoboman, A. E., 2017. Influence of power quality problem on the performance of an induction motor. American Journal of Electrical Power and Energy Systems, Vol: 4, No:4, pp: 39-44.

Bessous, N., Zouzou, S. E., Bentrah, W., Sbaa, S., and Sahraoui, M., 2018. Diagnosis of bearing defects in induction motors using discrete wavelet transform. International Journal of System Assurance Engineering and Management, Vol: 9, No: 2, pp: 335-343.

Chouhan, A., Gangsar, P., Porwal, R., Mechefske, C. K., 2021. Artificial neural network–based fault diagnosis for induction motors under similar, interpolated and extrapolated operating conditions. Noise & Vibration Worldwide, Vol: 52, No: 10, pp. 323-33.

Ding, Xiaoxi, and He, Q., 2017. Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. IEEE Transactions on Instrumentation and Measurement, Vol: 66, No: 8, pp: 1926-1935.

Dong, Shaojiang, Xu, X., and Chen, R., 2016. Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing. Journal of the Brazilian Society of Mechanical Sciences and Engineering, Vol: 38, No: 8, pp: 2255-2263.

Donolo, Pablo D., Pezzani, C. M., Bossio, G. R., Angelo, C. H. D., and Donolo, M. A., 2020. Derating of induction motors due to power quality issues considering the motor efficiency class. IEEE Transactions on Industry Applications, Vol: 56, No: 2, pp: 961-969.

Ghaseminezhad, Morteza, Doroudi, A., Hosseinian, S. H., and Jalilian, A., 2017. Analysis of voltage fluctuation impact on induction motors by an innovative equivalent circuit considering the speed changes. IET Generation, Transmission & Distribution, Vol: 11, No: 2, pp: 512-519.

Ghods, Amirhossein, and Lee, H., 2016. Probabilistic frequency-domain discrete wavelet transform for better detection of bearing faults in induction motors. Neurocomputing, Vol: 188, pp: 206-216.

Hassan, Ola., E., Amer, M., Ahmed, K., Abdelsalam, and Williams, B. W., 2018. Induction motor broken rotor bar fault detection techniques based on fault signature analysis–a review. IET Electric Power Applications, Vol: 12, No: 7, pp: 895-907

Hsueh, Yu-Min, Ittangihal, V. R., Wu, W., Chang, H., and Kuo, C., 2019. Fault diagnosis system for induction motors by CNN using empirical wavelet transform. Symmetry, Vol: 11, No: 10, pp: 1212.

Ibrahim, A., Anayi, F., Packianather, M., and Alomari, O. A., 2022. New hybrid invasive weed optimization and machine learning approach for fault detection. Energies, Vol: 15, No: 4, pp. 1488.

Kumar, Prashant, and Hati, A., S., 2021. Deep convolutional neural network based on adaptive gradient optimizer for fault detection in SCIM. ISA transactions, Vol: 111, pp: 350-359.

Kumar, R. R., Cirrincione, G., Cirrincione, M., Tortella, A., and Andriollo, M., 2021. Induction Machine Fault Detection and Classification Using Non-Parametric, Statistical-Frequency Features and Shallow Neural Networks. IEEE Transactions on Energy Conversion, Vol: 36, No: 2, pp. 1070-1080.

Liu, Yiqi, and Bazzi, A. M., 2017 A review and comparison of fault detection and diagnosis methods for squirrel-cage induction motors: State of the art. ISA transactions, Vol: 70, pp: 400-409.

Ma, Suliang, Chen, M., Wu, J., Wang, Y., Jia, B., and Jiang, Y., 2018. High-voltage circuit breaker fault diagnosis using a hybrid feature transformation approach based on random forest and stacked autoencoder. IEEE Transactions on Industrial Electronics, Vol: 66, No: 12, pp: 9777-9788.

Toma, N., Rafia, and Kim, J., 2020. Bearing fault classification of induction motors using discrete wavelet transform and ensemble machine learning algorithms. Applied Sciences, Vol: 10, No: 15, pp: 5251.

Pandarakone Esakimuthu, S., Mizuno, Y., and Nakamura, H., 2016. Distinct fault analysis of induction motor bearing using frequency spectrum determination and support vector machine. IEEE Transactions on Industry Applications, Vol: 53, No: 3, pp: 3049-3056.

Su, Jinya, and Chen, W., 2017. Model-based fault diagnosis system verification using reachability analysis. IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol: 49, No: 4: 742-751.

Toma Nishat, R., Prosvirin, A. E., and Kim, J., 2020. Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. Sensors, Vol: 20, No: 7, pp: 1884.

Xu, Xiaojian, Yan, X., Sheng, C., Yuan, C., Xu, D., and Yang, J., 2017. A belief rule-based expert system for fault diagnosis of marine diesel engines. IEEE transactions on systems, man, and cybernetics: systems, Vol: 50, No: 2, pp: 656-672.

Zaman, S. M. K., and Liang, X., 2021. An Effective Induction Motor Fault Diagnosis Approach Using Graph-Based Semi-Supervised Learning. IEEE Access, Vol: 9, pp. 7471-7482.

Zhang, Dongdong, An, R., and Wu, T., 2018. Effect of voltage unbalance and distortion on the loss characteristics of three-phase cage induction motor. IET Electric Power Applications, Vol: 12, No: 2, pp: 264-270.

Zhang, Nannan, Wu, L., Yang, J., and Guan, Y., 2018. Naive bayes bearing fault diagnosis based on enhanced independence of data. Sensors, Vol: 18, No: 2, pp: 463.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 3 (2022), 367-394
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.28821

IoT-Based Vision Techniques in Autonomous Driving: A Review

Mohammed Qader Kheder and Aree Ali Mohammed

Department of Computer, College of Science, University of Sulaimani, Iraq
✉ mohammed.kheder@univsul.edu.iq, aree.ali@univsul.edu.iq

KEYWORDS

autonomous driving; AVs; AVs' sensors; computer vision; Internet of Thing (IoT); Internet of Vehicles (IoV); traffic; accident prevention

ABSTRACT

As more people drive vehicles, there is a corresponding increase in the number of deaths and injuries that happen due to road traffic accidents. Thus, various solutions have been proposed to reduce the impact of accidents. One of the most popular solutions is autonomous driving, which involves a series of embedded systems. These embedded systems assist drivers by providing crucial information on the traffic environment or by acting to protect the vehicle occupants in particular situations or to aid driving. Autonomous driving has the capacity to improve transportation services dramatically. Given the successful use of visual technologies and the implementation of driver assistance systems in recent decades, vehicles are prepared to eliminate accidents, congestion, collisions, and pollution. In addition, the IoT is a state-of-the-art invention that will usher in the new age of the Internet by allowing different physical objects to connect without the need for human interaction. The accuracy with which the vehicle's environment is detected from static images or videos, as well as the IoT connections and data management, is critical to the success of autonomous driving. The main aim of this review article is to encapsulate the latest advances in vision strategies and IoT technologies for autonomous driving by analysing numerous publications from well-known databases.

1. Introduction

The most serious health problems are caused by traffic collisions. The number of deaths and injuries on the road generally increases as the number of vehicles on the road increases. Every year, about 1.3 million people die as a result of road traffic accidents, with another 20–50 million being disabled or injured (WHO, 2020). Engine industries have actively pursued new technology to increase passenger safety since the dawn of the automobile industry (Varma et al., 2019). Simple mechanical systems, such as turn signals or seat belts, gave way to more advanced ones, namely airbags or pop-up hoods, in the early stages. Later, as electronics progressed, many new developments emerged, for instance, electronic stability control and vehicle brake warning, which provided increased protection by signal processing (Zhao et al., 2021; Cheng et al., 2019). Moreover, some basic emerging systems that offer more passenger protection, such as driver assistance systems based on computer vision techniques, have arisen in the last decade, and several companies have implemented them in their vehicles (Agarwal et al., 2020; Sakhare et al., 2020; Varma et al., 2019).

During the last decade, applications built on mobile devices, sensors, actuators, and other devices, have become smarter, allowing for system connectivity and the implementation of more complex tasks. According to (Sharma, 2021; Litman, 2020; Celesti et al., 2018), the number of connected devices exceeded the global population a few years back, and the number has continued to rise exponentially until now. Wireless sensors, embedded systems, mobile phones, and nearly all other electronic items are now connected to a local network or the Internet, bringing with them the IoT era (Whitmore et al., 2015). The IoT has been around for a few decades and has gained popularity with the advancement of modern wireless technologies (Wang et al., 2021; Lu et al., 2019). Despite there being several different definitions of the IoT, in essence, IoT enables the virtual world to cross the boundaries of the physical world. In addition, the IoT can be described as a global network that connects uniquely specified virtual and physical things, smart objects, and devices, allowing them to communicate with one another (Wang et al., 2021; Fernando et al., 2020; Litman, 2020; Kang et al., 2018). What is more, the IoT paradigm is known as «any-place, any-time, any-one» connected (Litman, 2020). As reported by (Celesti et al., 2018), our daily environment, which includes vehicles, homes, streets, roads, or workplaces, may be connected to smart sensors, actuators, embedded systems, which can communicate with one another through a gateway device.

Nowadays, the IoT is a state-of-the-art technology that can be applied to a wide range of uses, including smart cities, smart transportation, and issues associated with deployment (Da Xu et al., 2014, Singh et al., 2014). A few examples of IoT services are shown in Figure 1, regarding robotics, intelligent transportation systems, smart cities, and healthcare. IoT is expected to connect billions of new physical devices to the Internet (Sharma, 2021; Wang et al., 2021; Minovski et al., 2020) and therefore, to grow services with a range of special properties and quality requirements. When all of the smart items connected to the Internet are automobiles, the IoT becomes the Internet of Vehicles (IoV). As a result, the IoV is an expanded application of the IoT for intelligent transportation. It is envisioned to be a key data sensing and processing platform for intelligent transportation systems (Kang et al., 2018).

Vehicles that interact with each other, as well as portable devices carried by roadside units, pedestrians, and the public network, use vehicle-to-vehicle (V2V), vehicle-to-human (V2H), vehicle-to-road (V2R), vehicle-to-device (V2D), vehicle-to-pedestrian (V2P), vehicle-to-infrastructure (V2I), vehicle-to-grid (V2G), and vehicle-to-sensor (V2S) interconnectivity to create a network in which the members are smart objects rather than humans (Fernando et al., 2020; Sadiku et al., 2018). The IoV environment is made up of the wireless network environment and road conditions. A vehicle might be a smart sensor platform that collects data from other vehicles, drivers, pedestrians, and the environment and uses it for safe navigation, traffic management, and pollution control (Sakhare et al., 2020). As can be seen, people use vehicles daily. The most serious concern regarding their expanding use is the increasing number of deaths that occur as a result of traffic accidents. The associated costs and risks are acknowledged as important issues confronting modern civilization (Celesti et al., 2018). For several years, researchers have been looking into Vehicular Ad-hoc Network (VANET), which evolved from

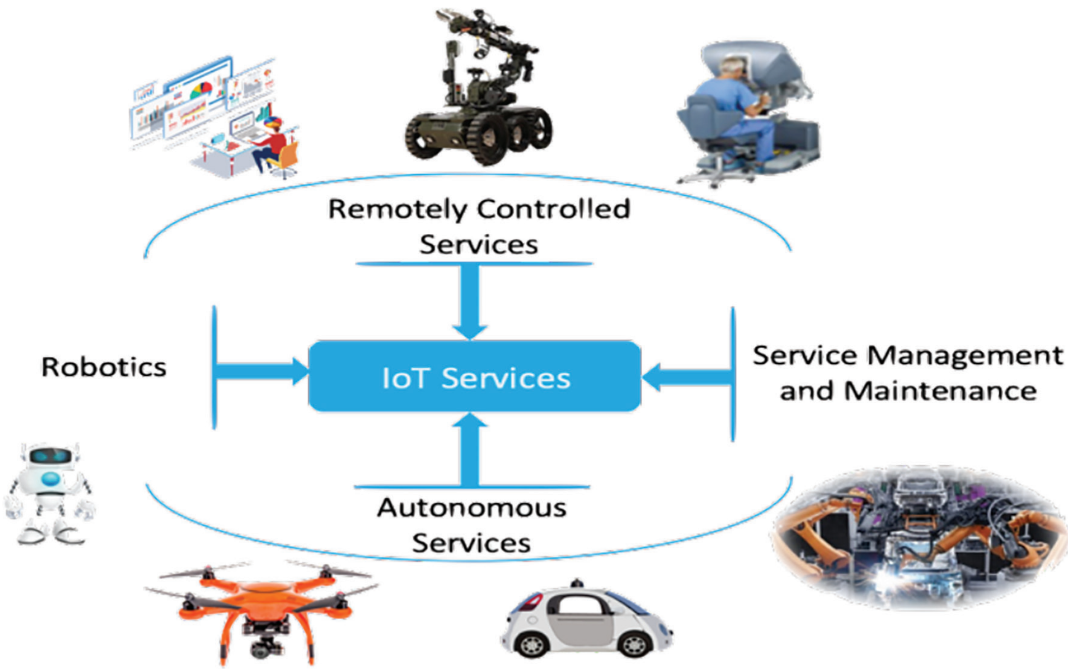


Figure 1. IoT services (Minovski et al., 2020)

Mobile Ad-hoc Network (MANET) (Sharma, 2021; Sadiku et al., 2018). Furthermore, because of people's lifestyle changes, a variety of VANET criteria are proposed. Many real-life scenarios need vehicle networking technology, such as driving on the highway. Drivers want to understand the traffic status on the roads ahead of them and modify their driving path if an accident or traffic congestion happens. There are many different ways that IoT technology and vision techniques are being used in the field of autonomous vehicles, and this review article looks at how these technologies and techniques are used in this field.

The rest of the paper is organized as follows. In Section 2, several theoretical concepts are discussed, including the term «IoT» and its structure, autonomous vehicle technology and various smart sensors, the use of the IoT for autonomous vehicles, and an explanation of how different vision techniques could be used to reduce road traffic accidents. Section 3 highlights some of the IoV challenges that may arise during the design and implementation of the IoV, as well as nine popular existing road vehicle systems that use IoT and computer vision. This research assessed a number of recent studies on the use of IoT technologies and vision methods in various autonomous vehicle applications, as well as compared the results of different autonomous systems employing various vision methods and algorithms, and also which methods should be used to achieve the best results in Section 4. Finally, Section 5 gives the conclusion and recommendations section.

2. Theoretical Concepts

2.1. The Internet of Thing and Its Structure

In 1999, Kevin Ashton proposed the concept of the «Internet of Things» (Hamid et al., 2020). The meaning of «things» has expanded over the last decade as technology has advanced. However, the key purpose remains the same: a digital computer can make sense of information without human interaction (Ravikumar and Kavitha, 2021; Fernando et al., 2020). In addition, according to (Singh et al., 2014), smart sensors, actuators, and an embedded system with a microprocessor are typical components of «things». They have to communicate with one another, building the need for M2M (Machine-to-Machine) communication (Wang et al., 2021; Ahmed and Pothuganti, 2020; Hamid et al., 2020). As reviewed by (Whitmore et al., 2015), the concept of the IoT is developing from M2M connectivity.

According to (Ravikumar and Kavitha, 2021; Fernando et al., 2020; Cheng et al., 2019), wireless technologies, such as Bluetooth, Wi-Fi, and ZigBee, can be used for short-range connectivity; however, mobile networks such as LoRa (Long Range), WiMAX (World Wide Interoperability for Microwave Access), GSM (Global System for Mobile), NB-IoT (NarrowBand-Internet of Things), Sigfox, CAT-M1 (Category M1), GPRS (General Packet Radio Service), 3G, 4G, LTE (Long Term Evaluation), and 5G are suitable for long-range communication. DeviceHive, Kaa, Thingspeak, Mainflux, and Thingsboard are IoT platforms that enable M2M communication via AMQP (Advanced Message Queuing Protocol), MQTT (MQ Telemetry Transport), STOMP (Simple Text Oriented Messaging Protocol), XMPP (Extensible Messaging and Presence Protocol), and HTTP (HyperText Transfer Protocol) (Cheng et al., 2019). These platforms provide node management, monitoring capabilities, data storage and analysis, and other features. Furthermore, based on their characteristics, these protocols may be employed for a variety of IoT applications. However, (Fernando et al., 2020; Cheng et al., 2019; Hanan, 2019) claim that the MQTT protocol is the most popular for IoT devices because of its features, which include the ability to gather data from a variety of electronic devices and provide remote device monitoring. It is also a TCP (Transmission Control Protocol) publish/subscribe protocol that is excellent for lightweight M2M communication over TCP with minimal data loss. Furthermore, the type of publish/subscribe protocol for the IoT reduces the need for clients to request updates, reducing network traffic and processing effort.

Moreover, since data processing occurs in the cloud computing infrastructure, it may sometimes be critical, depending on the application, that some data processing take place in IoT devices rather than a centralized node. Hence, as the processing partially moves to the end network elements, a modern computing model is introduced, which is called edge computing (Hanan, 2019). These devices are not suitable for handling intense processing tasks because they are mostly low-end devices. As a consequence, a fog node is required in the middle (Rani et al., 2021; Hanan, 2019). As stated by (Rani et al., 2021; Hanan, 2019; Kang et al., 2018), fog nodes have many characteristics, including sufficient resources, the ability to perform advanced computing tasks, the ability to be physically placed near end network components, and the ability to mitigate the overload generated by massive data transfer to any central cloud node. Furthermore, they assist IoT devices with big data handling by giving computing, storage, and networking services. Finally, the data is stored on a cloud server, where it is available for deep processing using a range of machine learning algorithms, as well as collaboration with other devices. Nowadays, many IoT applications are available for use. Table 1 discusses several common IoT applications:

Table 1. Common IoT applications

No.	Application's Name	Details
1	Smart Transportation	Smart physical sensors and actuators can be integrated into vehicles, as well as mobile applications and devices installed in the area. It is possible to have customized route recommendations (Elbery et al., 2020), economic street lighting (Sastry et al., 2020), quick parking reservations (Ahmed and Pothuganti, 2020), accident avoidance (Choubey and Verma, 2020; Celesti et al., 2018), automated driving (Rahul et al., 2020), and other services.
2	Smart Homes	Contains conventional devices such as washing machines, refrigerators, light bulbs, a smart door that opens and locks, and other appliances that have been built and can connect with licensed users through the Internet, allowing for improved system control and maintenance as well as energy usage optimization (Zaidan and Zaidan, 2020).
3	Security and Surveillance	Often, companies have monitoring and surveillance systems, and they have smart cameras that can collect video feedback from the road or a particular location. Smart detection technologies could be able to find criminals or avoid dangerous situations by recognizing objects in real time (Sultana and Wahid, 2019).
4	Health-Care Assistance	Smart devices for health-care assistance are being built to improve a patient's well-being. Plasters with remote sensors also track the status of a wound and report the data to experts without requiring physical examination (Philip et al., 2021; Banerjee et al., 2020). According to (Philip et al., 2021), sensors such as wearable devices or small inserts would be able to detect and monitor a wide range of parameters, including blood oxygen and sugar levels, heart rate, or temperature.
5	Environmental Conditions Monitoring	A network of remote sensors spread across the city provides the ideal infrastructure for monitoring a wide range of environmental conditions. Several weather sensors, such as barometers, ultrasonic wind, and humidity sensors, can assist in the construction of advanced weather stations (Lex et al., 2019). Several smart sensors are also able to track the city's air quality and water pollution levels (Srivastava and Kumar, 2021; Lex et al., 2019).

The IoT can be thought of as a vast network made up of subnetworks of connected devices and computers interconnected by some intermediaries, with various technologies such as Radio Frequency Identification (RFID), barcodes, and wired and wireless connections acting as network enablers (Sharma, 2021; Da Xu et al., 2014). The International Telecommunication Union (ITU) defines four dimensions of things as views of the IoT framework, as shown in Figure 2.

- **Tagging Things**- RFID is at the forefront of the IoT vision because of its real-time thing traceability and addressability. RFID serves as an automated barcode to better distinguish something that is connected automatically (Whitmore et al., 2015; Da Xu et al., 2014). Active RFID, which has a battery on board and is often used in healthcare, retail, and facilities management, and passive RFID, which does not have a battery and is powered by the reader and is widely used in road toll tags and bank cards, are the two varieties of RFID (Elbasani et al., 2020; Da Xu et al., 2014; Singh et al., 2014).

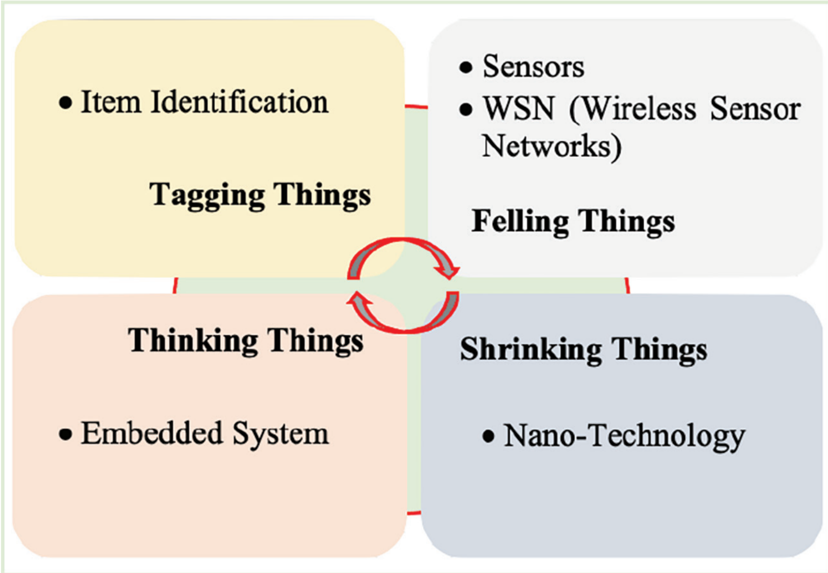


Figure 2. The 4th IoT dimensions (ITU)

- **Felling Things**- smart sensors and actuators serve as the key instruments for collecting data from the environment. Communications are built to connect the physical and information worlds, so that data may be obtained (Hirz and Walzel, 2018). Recent technical advances have resulted in high performance, low power consumption, and low cost.
- **Shrinking Things**- nanotechnology and miniaturization have enabled smaller things to communicate in things or smart devices. The key benefit is an increase in quality of life, such as the use of nano-sensors to track water quality at a lower rate (Shashank et al., 2020), as well as in healthcare (Li et al., 2021; Whitmore et al., 2015), for example, the use of nano-sensors in cancer detection and treatment.
- **Thinking Things**- embedded intelligence in devices via smart sensors has formed the network connection to the Internet. Local electric applications can be developed to achieve intelligent control, such as refrigerators that can detect the quantity of different items and the freshness of perishable items. It may also be possible to communicate with clients by sending them warnings over the Internet.

2.2. Autonomous Vehicles and Sensors

Autonomous Vehicles (AV): It is a vehicle that can operate without human intervention (Hamid et al., 2020; Litman, 2020). It is any vehicle that has features that allow it to accelerate, brake, and steer without the need for human intervention. According to (Ravikumar and Kavitha, 2021; Hamid et al., 2020; Litman, 2020), the concept of AVs dates back to the 1930s, when science fiction visualized and innovated AVs as a new challenge for the automotive industry. Several leading automotive manufacturers around the world, including Audi, Ford, Volvo, Nissan, and BMW, have started to build autonomous vehicles (Ahmed and Pothuganti, 2020; Lu et al., 2019). Figure 3 depicts the evolution of AVs from 1998 to 2018.

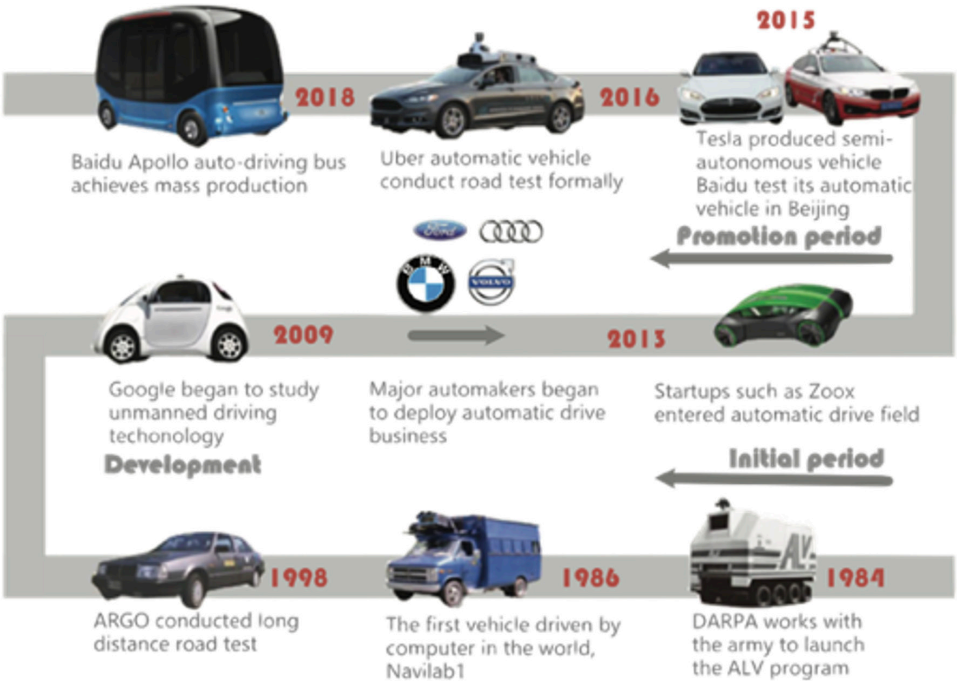


Figure 3. The development of AVs from 1998 to 2018 (Lu et al., 2019)

The AVs use various technologies and sensors, including adaptive cruise control, anti-lock brakes (brake by wire), active steering (steer by wire), lasers, radar, and GPS navigation technology (Ravikumar and Kavitha, 2021; Hamid et al., 2020; Litman, 2020; Lu et al., 2019). As a result, AVs have been targeted for their ability to reduce pollution and dust particles, prevent accidents, improve vehicle safety, reduce fuel consumption, create new potential market opportunities, and free up driver time (Ravikumar and Kavitha, 2021; Ahmed and Pothuganti, 2020). Figure 4 represents the AV system's block diagram, which is divided into four categories (Kocić et al., 2018).

According to the estimations of (Ahmed and Pothuganti, 2020; Lu et al., 2019), nearly 10 million autonomous vehicles will be on the road in the next few years. (Ahmed and Pothuganti, 2020; Litman, 2020; Lu et al., 2019) divides two categories of AVs: semi-autonomous and fully-autonomous, as described below:

1 *Semi-Autonomous*- can accelerate, steer and brake, keeping the distance from the car in front and also keeping the lane speed up. However, the driver is still needed and is still in full control.
2 *Fully-Autonomous*- capable of driving from a starting point to a fixed destination without the driver's intervention. This form of AV is expected to debut in the coming years.

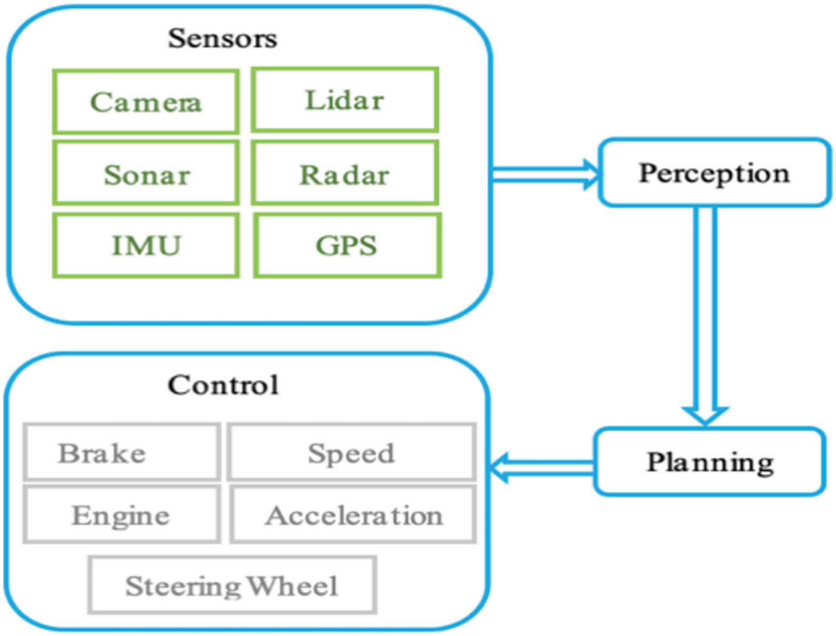


Figure 4. The block diagram of the AV's system (Kocić et al., 2018)

AVs can have several benefits; however, they may also entail many costs: internal (affecting the driver) and external (affecting other people). As a result, when preparing for autonomous vehicles, all of these factors should be taken into account (Ravikumar and Kavitha, 2021; Litman, 2020). The table below shows the possible benefits and drawbacks of autonomous vehicles, which should be carefully thought about before they are used in a project.

Table 2. AVs benefits and drawbacks (Litman, 2020)

	Benefits	**Drawbacks**
Internal	The stress level of drivers has decreased, and their productivity has increased.	Vehicle prices have gone up because more equipment, services, and taxes have been added to vehicles.
	More independent mobility for non-drivers can minimize chauffeuring responsibilities and transit subsidy demands for motorists.	More accidents are caused by increased user hazards, including system failures, platooning, higher traffic speeds, increased risk-taking, and increased overall vehicle travel.
	Reduced paid driver costs- this lowers the cost of taxis and commercial transportation drivers.	Features such as location tracking and sharing data may make it hard to keep your data safe and private.

Table 2. AVs benefits and drawbacks (Litman, 2020) (continued)

	Benefits	Drawbacks
External	Increased safety may reduce the likelihood of a collision and insurance costs.	Increased infrastructure costs may necessitate stricter standards for roadway construction and maintenance.
	Increased road capacity and cost savings from more efficient traffic could cut down on congestion and road costs.	Additional dangers- situations that may endanger other road users and be used for illegal purposes.
	Parking costs are reduced, which means less demand for parking at locations.	Congestion, pollution, and sprawl-related expenses may all rise as a result of better automobile travel.
	Reduced energy usage and pollutants might help to improve fuel economy and reduce pollution.	Concerns about social equality- walking, biking, and public transportation would continue to be the cheapest ways to get around.

According to the SAE (Society of Automotive Engineers) organization, vehicles are classified into six levels in terms of autonomous vehicles, as indicated in Figure 5, starting at level zero and ending at level five (Hamid et al., 2020).

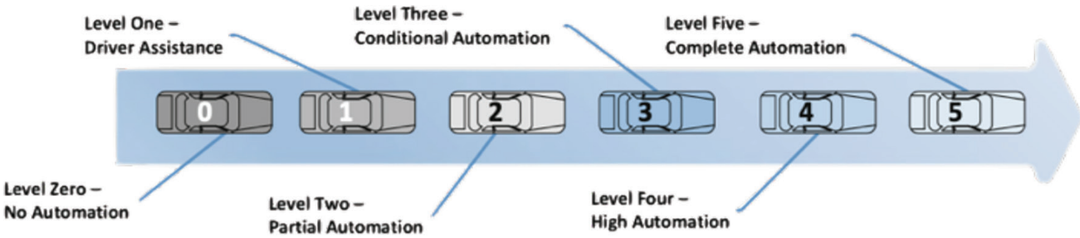


Figure 5. The journey of automation to fully AVs (Hamid et al., 2020)

At Level 0 (no automation), people do all the complicated driving operations such as accelerating or slowing down, steering, braking, and so forth. Level 1 (driver assistance): acceleration or deceleration aid, or steering based on knowledge of driving circumstances. Level 2 (partial automation): automatic acceleration and deceleration, as well as steering, are combined. Level 3 (conditional automation): a driving mode in which an automated driving system reacts exactly to a request when the driver replies. Level 4 (high automation): In some cases, even if a person does not reply to a request, vehicle automation can perform all driving tasks. Level 5 (full automation): the whole automation system of the vehicle is capable of performing all driving activities in any circumstances (Hamid et al., 2020). By using a lot of different devices and smart sensors and actuators to run and manage a driverless car at Levels 4 and 5, AVs can do all of the things that people do when they drive.

Sensors: As claimed by (Hamid et al., 2020; Hirz and Walzel, 2018; Da Xu et al., 2014), sensors are the most important components of AVs. Thus, AV would be impossible without sensors, which allow the vehicle to see and feel everything on the road while also gathering the necessary data to drive safely. There are several smart sensors: video cameras, lidar, radar, sonar, a global positioning system (GPS), the inertial measurement unit (IMU), wheel odometry, and other smart sensors and actuators that can all be part of autonomous vehicles (Kocić et al., 2018). Sensors can gather data, which is subsequently analyzed by the AV's computer and utilized to monitor the vehicle's steering, speed, and braking (Hamid et al., 2020; Kocić et al., 2018). More to say, data collected by sensors in AVs, such as traffic jams, the actual path ahead, and any road obstacles, can be exchanged among vehicles linked by M2M technology, a process known as V2V (vehicle-to-vehicle) communication (Singh et al., 2014). Several AVs' sensors are covered in the table below and references are provided to the reviewed articles:

Table 3. Common sensors in autonomous vehicles

No.	Article(s)	Sensor	Sensor Details
1	(Cao et al., 2019; Koresh and Deva, 2019; Minhas et al., 2019;)	Camera	Detects real-time obstacles to allow for lane departure and road sign information. The captured image contains a large number of values for individual pixels; however, these figures are almost useless. Therefore, computer vision algorithms should be used to convert low-level image information into high-level image information.
2	(Ahmed and Pothuganti, 2020; Xiang et al., 2018)	Ultrasonic	It is typically used to detect obstacles over small distances, for example, to assist drivers with parallel parking or using auto-parking devices. It also uses a high-frequency sound wave that bounces back to measure the objective distance of a vehicle. Ultrasonic technology sends out sound waves at a frequency of 50 kHz and waits for a response.
3	(Khalifa et al., 2020; Dhawale and Gavankar, 2019)	Radio Detection and Ranging (RADAR)	RADAR emits radio waves that detect depth over short and long distances. They are dotted around the vehicle to monitor the position of the vehicle nearby. By calculating the difference in frequency of returned signals generated by the doppler effect, this sensor can detect the distance and speed of moving objects. In addition, it is built into vehicles for a number of different reasons, such as adaptive cruise control, blind-spot detection, and collision avoidance or warning.
4	(Khalifa et al., 2020; Kocić et al., 2018)	Light Detection and Ranging (Lidar)	Lidar determines the distance between the sensor and the nearest object using an infrared laser beam. It uses active pluses of light from the vehicle's surroundings to sense route boundaries and distinguish lane markers. The most modern lidar sensors use light with a wavelength of 900 nm, while some lidars use longer wavelengths that perform better in rain and fog.
5	(Hamid et al., 2020)	Dedicated Short-Range Communications (DSRC)	DSRC is a collection of standards and protocols for one-way or two-way short-range to medium-range wireless communication networks specifically intended for automotive vehicle use. It can be used for 4G, Wi-Fi, Bluetooth, and others for V2V, V2I (Vehicle-to-Infrastructure) communication, and V2X (Vehicle-to-Everything).

Table 3. Common sensors in autonomous vehicles (continued)

No.	Article(s)	Sensor	Sensor Details
6	(Choubey and Verma, 2020; Elbery et al., 2020; Celesti et al., 2018)	Global Positioning System (GPS)	GPS receivers allow autonomous vehicles to maneuver without the need for human input. AVs can automatically sense, store, and retrieve data about their surroundings while other technologies such as laser light, radar, and computer vision are integrated. However, more sophisticated techniques are required for autonomous vehicles to execute certain activities that people can perform, such as parking (Ahmed and Pothuganti, 2020), recognizing road signs (Choubey and Verma, 2020), and detecting diverse vehicles (Xiang et al., 2018) on the road.

2.3. IoT in Autonomous Driving

The term «automotive IoT» refers to embedding IoT technology into automotive systems to create modern applications that make the vehicle more intelligent, more powerful, and safer to drive (Celesti et al., 2018; Sadiku et al., 2018). The concept of IoT in the automotive industry could be summarized into three main divisions: connectivity among vehicles, which is related to connectivity from one vehicle to another (V2V), connectivity between vehicles with external infrastructure (V2I), and connectivity with external hardware and devices (V2D). Taking this concept further, a connected vehicle is a vehicle that includes a platform which allows for the exchange of information between the vehicle and its surroundings using local wireless, the Internet, or smart sensors (Minovski et al., 2020). Wang et al., 2021, state that a typical IoT platform is an integrated system capable of supporting hundreds of millions of simultaneous device communications to generate large amount of data to be transferred and processed in cloud computing. In addition, according to (Fernando et al., 2020; Hamid et al., 2020; Rahul et al., 2020), a typical IoT platform contains four main components, as shown in Figure 6, all of which are concerned with the IoT autonomous vehicle platform: Smart sensors and hardware devices are important components that collect various data types from the real world, as well as a communication network component that is typically built on wireless technology, namely cellular technologies (3G, 4G, and 5G) or Wi-Fi, and a big data component that represents the amount of data being generated; this data should be transmitted, stored, and processed; and the cloud component where the data will be stored and processed because it provides some processing, analytics, and storage services. As reviewed by Da Xu et al., 2014, traditional IoT applications are hosted in the cloud and can provide input and make choices for physical systems. In IoT AVs, on the other hand, the cloud will serve as the centralized administration system for all software components and monitoring tools (Hamid et al., 2020; Celesti et al., 2018).

IoT for AVs is transforming the transportation system into a global variety of vehicular networks. Smart vehicle monitoring, dynamic information systems, and software to minimize insurance costs, road congestion, and potential collisions are just a few of the advantages of the IoT. There are two different data-gaining modules in the IoT AV platform. To begin with, collect data from personal sensors and share it with others in the neighbourhood. Second, the IoT platform serves as a repository for massive amounts of data received from various gateways (e.g., traffic information, environmental information, parking, and transportation) and connected devices (e.g., rail, traffic lights, a car, working areas, parking spots, entertainment, pedestrians, weather conditions, other vehicles, and so on) (Choubey and Verma, 2020).

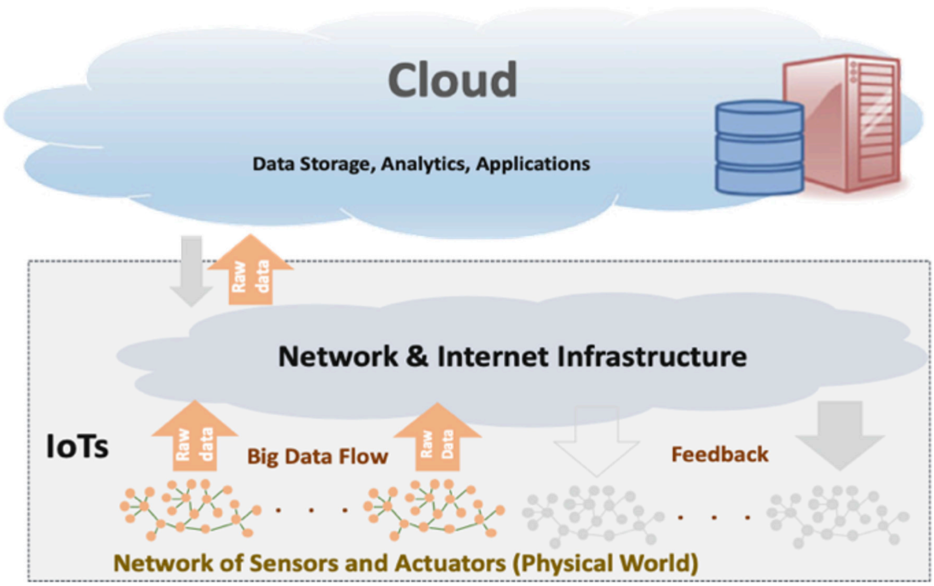


Figure 6. Typical components of the IoT platform (Hamid et al., 2020)

2.4. Road Traffic Accident Prevention using Vision Techniques

Road traffic accidents, which are one of the leading causes of injuries and deaths in the world, have been exacerbated by the rapid rise in the number of vehicles on the road. As a result, these numbers will cause more and more traffic accidents on the road (Uma and Eswari, 2021; Agarwal et al., 2020). According to (Uma and Eswari, 2021; Agarwal et al., 2020; Choubey and Verma, 2020), the trigger factors are classified into two categories: human and environmental, with the human aspect being far more responsible for the number of accidents. As reported by the National Highway Traffic Safety Administration (NHTSA), 85 percent of all road accidents happened in developing countries. Moreover, (Agarwal et al., 2020; Celesti et al., 2018) discuss that a lack of attention due to driver fatigue or drowsiness was a factor in one out of every three crashes in 2019, while drugs, alcohol, or drunk driving, and interruption were factors in one out of every five accidents. As claimed by (Uma and Eswari, 2021; Celesti et al., 2018) the following are the most frequent causes of traffic accidents:

- *Frontal crashes* occur when two vehicles traveling in opposite directions intersect. Although different types of mechanisms exist that can reduce the number of injuries that a frontal vehicle collision can cause, it may be seen that computer vision strategies have the ability to prevent a potential or immediate automotive collision. Using one or more frontal video cameras, the vehicle would be able to identify a potentially risky situation and respond appropriately, such as with an alert or sound (Uma and Eswari, 2021; Agarwal et al., 2020). Moreover, as claimed by (Hamid et al., 2020; Minovski et al., 2020), a variety of sensors, including lidar and radar, can assist cars in avoiding frontal collisions by monitoring the environment around them.

- *Lane departure collisions* occur when a driver changes lanes, causing the vehicle to crash with another vehicle in a similar lane. (Dhawale and Gavankar, 2019; Narote et al., 2018; Zang et al., 2018) state that the computer vision techniques greatly reduce the number of injuries caused by lane departure collisions. If the driver is distracted, the system can effectively complement the risky behaviour in similar ways, such as changing lanes and focusing the driver's focus. Furthermore, since the device is installed in the car, drivers are more careful to avoid crossing the lane marker as a result of any form of warning, such as an audible alert, a vibrating driver seat, a graphical note, and others (Olanrewaju et al., 2019; Zang et al., 2018). The following figure 7 is a block diagram of a lane departure collision.

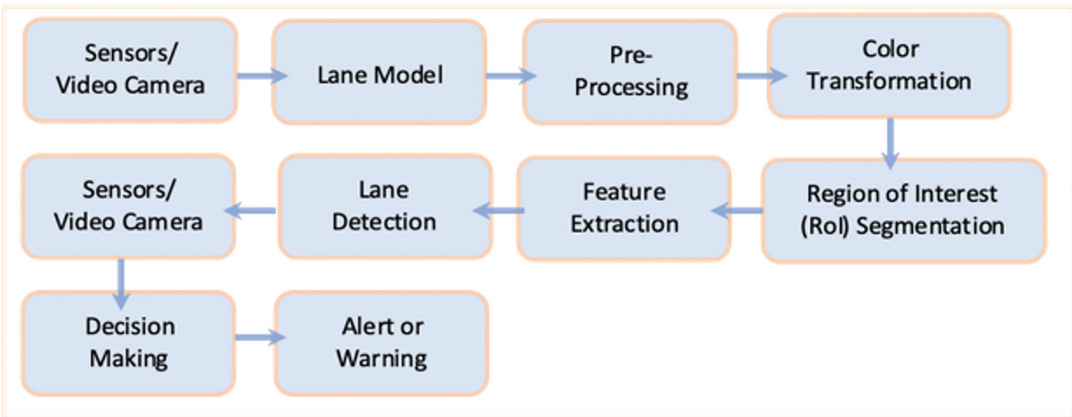


Figure 7. The block diagram of lane departure collision (Dhawale and Gavankar, 2019; Zang et al., 2018)

- *Surrounding vehicles*: If the drivers are not aware of them, they can be dangerous. A driver's capacity to identify surrounding vehicles is considerably reduced when he/she is tired or fatigued. This can lead to a collision in situations such as driving in a convey, changing lanes, parking, and other similar situations (Agarwal et al., 2020). Depending on the complexity of the system that monitors the surrounding vehicles, appropriate action, such as sending an alert, is taken, depending on the actual vehicle condition and the situation anticipated in the immediate future. As believed by (Chetouane et al., 2020; Xiang et al., 2018), the main issue with this vision method is the incorrect estimation of close vehicles. Since vehicles come in a variety of colors and sizes, this can lead to inaccurate estimates and unsafe driver behavior, endangering other drivers and road users.
- *Road signalization*: The key factor in traffic safety is road signalization. Drivers must be able to identify and detect traffic signals. The frequency of casualties caused by this traffic problem can be reduced if a suitable road signalization detection and a driver assistance device are employed (Ciuntu and Ferdowsi, 2020; Li et al., 2020; Narote et al., 2018). The technology notifies the driver if they make a potentially dangerous movement.

3. Study Conducted

In this section, the study discusses many IoV challenges that should be avoided while doing research on AVs, as well as some popular current AV applications.

3.1. The Internet of Vehicles' Challenges

While, after many years of research and development, AVs have become a reality, there are still many challenges in fully developing autonomous systems for engineering technologies, regulatory difficulties, a lack of industry-standard technology and equipment, consumer confidence and acceptance, and so on. These are all examples of AVs (Hamid et al., 2020; Whitmore et al., 2015, Singh et al., 2014). The challenges get more complex as the degree of autonomy increases. Security and privacy, big data, reliability, road conditions or environments, mobility, and open standards are among the problems facing IoV and delaying its acceptance. These challenges ought to be addressed for IoV to become very reliable and broadly accepted (Hamid et al., 2020; Sadiku et al., 2018).

- *Security* and *privacy* are essential because the IoVs need the integration of several different systems, standards, and services (Garg et al., 2020; Whitmore et al., 2015). IoV can be a target for cyber-attacks and intrusions because it is an open, public network. This can lead to physical and personal information being leaked (Whitmore et al., 2015).
- *Big Data*-due to the huge number of linked vehicles, one of the most difficult challenges of the IoV is the collection and storing of big data (Garg et al., 2020; Sadiku et al., 2018). As stated by (Hamid et al., 2020) four data aspects are explored in AVs: data collection, data handling, data management, and data labelling. In order to handle big data, big data analytics, and mobile cloud computing, it would be crucial.
- *Reliability vehicles*, smart sensors and actuators, and network hardware all have the potential to fail. The system should be able to handle erroneous data as well as unreliable communications, such as DoS attacks. In general, vehicle protection takes precedence over entertainment (Choubey and Verma, 2020; Sadiku et al., 2018).
- *Road Condition*-the state of the roads is highly variable and varies greatly from one location to the next (Ng et al., 2019; Narote et al., 2018). For example, some areas have smooth and well-marked large highways, while others have deteriorated road conditions with no lane marking (Hamid et al., 2020). Lanes are not well marked, there are potholes, and there are mountainous and tunnel roads with a lack of well-defined external path signals (Dhawale and Gavankar, 2019).
- *Mobility*-in a situation where vehicles move quickly and network topology changes often, it can be difficult to keep nodes connected and provide them with the services they need to send and receive data in real-time (Sadiku et al., 2018).
- *Open Standards*-interoperability and standardization are needed to drive adoption. According to (Garg et al., 2020; Sadiku et al., 2018), a lack of standards makes successful V2V collaboration impossible, although implementation of open standards facilitates sharing of information. In addition, governments should work with businesses to implement technical best practices and open international standards, as well as encourage them to work together, too.

- There are also many recent AV problems, such as weather conditions, traffic conditions, accident liability, radar interference, vehicular communication, and others, all of which affect the AV's efficiency.

3.2. Existing Road Vehicle Systems

In today's world, there are lots of initiatives utilizing computer vision methods and IoT in road vehicles. The major purpose of these initiatives is to build driver state recognition, and road infrastructure recognition modules for autonomous vehicles (Sharma, 2021; Ahmed and Pothuganti, 2020; Litman, 2020). According to (Litman, 2020; Lex et al., 2019), using IoT applications in the automotive industry aids in design by increasing performance, lowering costs, and allowing for quality control. Table 4 contains a list of existing systems that have been evaluated in a variety of publications; however, more recent articles have been selected.

Table 4. Existing autonomous vehicles' systems

No.	**Papers (s)**	**Proposed System's Name**	**Proposed System's Details**
1	(Choubey and Verma, 2020; Celesti et al., 2018)	Accident Detection/ Prevention	The main aim of this AV system is to detect accident-prone areas to minimize the number of accidents and traffic congestion. It can also notify drivers of crucial situations, allowing them to take action quickly.
2	(Ciuntu and Ferdowsi, 2020; Li et al., 2020; Zhang et al., 2020; Cao et al., 2019; Koresh and Deva, 2019)	Real-time Traffic Sign Detection	Traffic-sign recognition aims to identify lots of types of traffic signs put on the road, such as speed limits, turn ahead, school, children, and others that assist drivers to keep safe driving. To detect and recognize distinct traffic signs, this type of system uses image processing algorithms. Furthermore, traffic sign detection may be achieved in two stages: color segmentation and classification.
3	(Liu et al., 2020; Savaş and Becerikli, 2020; Minhas et al., 2019; Eddie et al., 2018; Izquierdo et al., 2018; Kumar and Patra, 2018)	Driver Fatigue Detection	This system is used to detect oncoming driver fatigue, and the issue of timely warning can aid in avoiding lots of accidents, and consequently reducing personal suffering. According to international statistics, driver fatigue causes a large number of road accidents; therefore, it plays a big role in AVs.
4	(Ahmed and Pothuganti, 2020; Chetouane et al., 2020; Wei et al., 2019; Hsu et al., 2018; Tsai et al., 2018; Xiang et al., 2018)	Vehicle Detection	There are a variety of autonomous sensors that may be used to identify nearby vehicles. This would allow drivers to be aware of what is going on around them. The optical camera is used in many types of research, which has a challenge due to vehicle shape, size, color, and environmental appearance. Knowledge-based, stereo-vision, and motion-based vision approaches are all employed to identify vehicles.

Table 4. Existing autonomous vehicles' systems (continued)

No.	Papers (s)	Proposed System's Name	Proposed System's Details
5	(Khalifa et al., 2020; Kim et al., 2020; Pranav and Manikandan, 2020; Preethaa and Sabari, 2020; Zahid et al., 2019; Qu et al., 2018)	Pedestrian Detection	Detecting objects in an image that are people has a long history; pedestrian identification has advanced significantly in the previous decade. To assure safe behavior in certain environments, autonomous vehicles must be able to detect and avoid pedestrians. However, automatically identifying human forms using a visual system is an extremely advanced process. Accurate pedestrian detection might have an immediate effect on vehicle safety or other applications.
6	(Dhawale and Gavankar, 2019)	Lane Detection and Departure Warning	Every autonomous vehicle requires lane recognition and departure warning, and it's very beneficial for the driver's assistance. Due to the various difficulties that this type of system faces, such as limited visibility of the lane line, vagueness of lane patterns, shadows, brightness, light reflection, and others, it is a very difficult issue. Lane detection is achieved via two steps: image enhancement and edge detection; and extraction of lane characteristics; and a calculation of road form from the processed image.
7	(Ng et al., 2019)	Road Anomalies Detection	Anomalies on the road cause discomfort to drivers and riders, and they can lead to mechanical failure or even injuries. Since road conditions have an immediate effect on many facets of transportation, road anomaly detection plays an important role in smart transportation. The main goal of this system is to identify bumps and potholes on the road and warn drivers, thus reducing vehicles injuries, road accidents, and traffic congestion.
8	(Ahmed and Pothuganti, 2020)	Smart Parking	Smart parking applications are developed to efficiently track available parking lots and provide drivers with a variety of reservation options. It uses an IoT-based system that sends data about available and occupied parking spots through the web or mobile applications. Several studies have used image data to detect free parking spots in large numbers, using a variety of machine learning algorithms. Each parking spot contains several IoT devices, such as smart sensors and microcontrollers, which are used to detect the presence of a car in a parking space and send the data to a centralized system. The drivers receive a real-time report on the availability of all parking spots and choose the best one.
9	(Elbery et al., 2020)	Route Optimization and Navigation	In cities, traffic congestion is a frequent problem that is only growing worse as the number of vehicles increases. To decrease traffic congestion, a route optimization and navigation program may be used to suggest the optimal path for a certain location. Both vehicle emissions and travel time are reduced as traffic congestion is reduced. This system requires all necessary information to be contained, such as the number and location of all necessary stops along the route, as well as delivery time windows.

4. Research Findings and Discussion

4.1. Literature Survey on Various IoT Technologies used in Autonomous Driving

This section of the study contains details on a variety of ongoing IoT research projects in AVs. They employed a range of IoT technologies to tackle specific challenges, as seen in Table 5.

Table 5. Survey on different IoT technologies used in AVs

No.	**Paper(s)**	**System's Name**	**Used Technology(s)**	**Pros** / **Cons**
1	(Choubey and Verma, 2020; Celesti et al., 2018)	Accident Detection/ Prevention	• GPS, GPRS, GSM, GPS TK103 and 4G • Arduino Mega 2560 • Ultrasonic Sensor • LM35 Temperature • ADXL335 Acc	**Pros:** • The accident rate can be reduced • Notify drivers while detecting an obstacle **Cons:** • Require 100% reliability and low maintenance
2	(Ciuntu and Ferdowsi, 2020; Li et al., 2020; Zhang et al., 2020; Cao et al., 2019; Koresh and Deva, 2019)	Real-time Traffic Sign Detection	• Video Camera	**Pros:** • Improve the driver's road safety, especially when he/she is tired or has missed signs • It is crucial when you are driving somewhere you have never been before **Cons:** • During busy hours, they cause traffic congestion by stopping vehicles at crossroads • During peak hours, when signals fail, there are significant and widespread traffic problems
3	(Liu et al., 2020; Savaş and Becerikli, 2020; Minhas et al., 2019; Eddie et al., 2018; Izquierdo et al., 2018; Kumar and Patra, 2018)	Driver Fatigue Detection	• CCD Micro Camera with Infrared Illuminator • CCD Camera • Digital Video Camera • IR Camera	**Pros:** • The detection of inappropriate behaviour is solved in real-time through alerts • The driver's life can be protected if he or she is alerted while driving in a dangerous situation **Cons:** • Accidents can be reduced to preserve traffic management • Due to multiple sensors located near the driver's eyes, the driver's body is suffering from health problems

Table 5. Survey on different IoT technologies used in AVs (continued)

No.	Paper(s)	System's Name	Used Technology(s)	Pros	Cons
4	(Ahmed and Pothuganti, 2020; Chetouane et al., 2020; Wei et al., 2019; Hsu et al., 2018; Tsai et al., 2018; Xiang et al., 2018)	Vehicle Detection	• Raspberry pi B+ • PI CAM • Ultrasonic sensor (HC-SR07) • UAV Technology • LCD Display	• Detect and identify objects that surround the vehicle, giving drivers an excellent sense of his/ her surroundings	• Vehicles come in a variety of shapes and colors. This can lead to inaccurate vehicle estimates and incorrect driver behaviours, putting other drivers and road users at risk
5	(Kim et al., 2020; Khalifa et al., 2020; Pranav and Manikandan, 2020; Preethaa and Sabari, 2020; Zahid et al., 2019; Qu et al., 2018)	Pedestrian Detection	• Radar • Lidar • Optical Camera • Infrared Radiation	• It aids the driver's ability to the drive safely • Stay safe pedestrian's life while walking on the road	• Traffic jams are caused by failed pedestrian detection
6	(Dhawale and Gavankar, 2019)	Lane Detection and Departure Warning	• Color Detection Sensor • Radar • Optical Camera	• Minimizing the number of auto accidents	• Vehicles are causing problems for vehicles around it
7	(Ng et al., 2019)	Road Anomalies Detection	• IoT Sensors • Arduino Uno • MPU6050 3-axis Accelerometer • Optical Camera	• Avoid vehicle accident • Protect vehicle engineering • Monitoring the road while driving	• You navigate the dashboard, receive reports, and receive alerts to make change
8	(Ahmed and Pothuganti, 2020)	Smart Parking	• Ultrasonic Sensor • Web Service • FIWARE • Mobile Phones/ Apps	• Reduces cruising time • Smart parking as a solution to traffic woes	• It is not recommended for high peak hour volume facilities • Requires a maintenance
9	(Elbery et al., 2020)	Route Optimization and Navigation	• GPS • Smart Phones • Google Map	• Improve customer satisfaction • Reduce operational costs • Save time in the scheduling of transportation	• Missing traffic data due to detector/ communication failure

4.2. The Comparison of Several Autonomous Systems' Results using Various Vision Techniques

One of the most significant characteristics of the system's effectiveness is the use of visual techniques. As a result, this review article summarizes the findings of several studies that used vision techniques in their proposed systems. Some of the above-mentioned systems, such as driver fatigue detection, real-time traffic sign detection, vehicle detection, pedestrian detection, and lane detection and departure, have been chosen since they are the most often used in AVs daily, and also several review articles have been chosen due to the best outcomes. In the following tables, the results of these systems are displayed. Then, recommendations on the best techniques or methods to use for each system are given. In addition, Confusion Matrix, F1-score, PERCLOSE (Percentage of Eye Closure), LogLoss (Logarithm Loss), IoU (Intersection over Union), AUC (Area Under Curve), ROC (Receiver Operating Characteristic), MAE (Mean Absolute Error), MSE (Mean Squared Error), MRR (Mean Reciprocal Rank), DCG (Discounted Cumulative Gain), NDCG (Normalized Discounted Cumulative Gain), etc., are some of the metrics that can be used to evaluate the proposed method's performance. Because each of them has its own math equation and work, the researchers should carefully pick one or a few of them to assess their systems. Also, some systems will produce the best outcomes for some of them while being bad for others.

Table 6-a. Outcomes of ***driver fatigue detection*** *system*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
1	(Izquierdo et al., 2018)	2018	• Electroencephalogram (EEG) • Camera • FACET Model	PERCLOSE and F1-score	• Gaussian Mixture Model	89%
2	(Eddie et al., 2018)	2018	• Percentage of Eye Closure (PERCLOS) • Artificial Vision	PERCLOSE	• Viola Jones Algorithm • Support Vector Machine (SVM)	93.3%
3	(Kumar and Patra, 2018)	2019	• Facial Landmark Marking • Head Bending • Yawn Detection	Confusion Matrix	• Histogram of Oriented Gradient (HOG) • Linear SVM	Bayesian Classifier: 85% FLDA: 92% SVM: 95%
4	(Minhas et al., 2019)	2019	• Steering Pattern Recognition • Vehicle Position in the Lane Monitoring • Driver's Eye or Face Monitoring • Physiology Measurement	Not Available	Convolutional Neural Network (CNN)	98.5%

*Table 6-a. Outcomes of **driver fatigue detection** system (continued)*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
5	(Savaş and Becerikli, 2020)	2020	• PERCLOS • Yawning Frequency/ Frequency of Mouth (FOM)	PERCLOSE	• Multi-Task ConNN	98.81%

According to the results, combining PERCLOS with FOM techniques and the ConNN deep learning algorithm produces the best results when compared to the alternatives.

*Table 6-b. Results of **real-time traffic sign detection** system*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
1	(Cao et al., 2019)	2019	• Color Space for Traffic Sign Segmentation • Shape Features for Traffic Sign Detection	Confusion Matrix	• LeNet-5 CNN	99.75%
2	(Koresh and Deva, 2019)	2019	• Color-based Segmentation • Hough Transform	Not Available	• Speeded up Robust Features (SURF)	57%
					• Scale-Invariant Feature Transform (SIFT)	40%
					• Binary Robust Invariant Scalable Keypoints (BRISK)	80%
					• CNN	88%
					• Capsule Neural Network (CapsNet)	98.2%
3	(Ciuntu and Ferdowsi, 2020)	2020	• Onboard Camera • Optical Character Recognition (OCR)	Confusion Matrix	• Mask R-CNN	93.1%
4	(Li et al., 2020)	2020	• Dynamic Threshold Segmentation (DTS) • Location Detection (LD) • Region of Interest (ROI)	Confusion Matrix and IoU	• SVM based on HOG	97.41%
5	(Zhang et al., 2020)	2020	• Dot-Product and SoftMax • Feature Pyramids • RoI	Not Available	• Multiscale Cascaded R-CNN	99.7%

The findings demonstrate that the LeNet-5 CNN algorithm, which uses color space for traffic sign segmentation and shape features for traffic sign detection, achieves the best results.

Table 6-c. Outcomes of ***vehicle detection*** *system*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
1	(Tsai et al., 2018)	2018	• Concatenated ReLU (C. ReLU) • Modified Inception • Hypernet • PVANET (Technique Used)	Not Available	• CNN	Over 90.3%
2	(Hsu et al., 2018)	2018	• Cameras • RoI	Confusion Matrix and IoU	• Fast R-CNN	88.085%
3	(Wei et al., 2019)	2019	• RoI • Precise Extraction	Confusion Matrix and ROC Curve	• Combined Harr Classifier with HOG • AdaBoost • SVM Classification	97.96%
4	(Ahmed and Pothuganti, 2020)	2020	• R-CNN Architectures • Single Shot Detectors (SSD) • You Only Look Once (YOLO)	Not Available	• Fast R-CNN	99.6%
5	(Chetouane et al., 2020)	2020	• Motion-based Approaches • Object-based Approaches	Confusion Matrix	• Gaussian Mixture Model (GMM)	75%
					• GMM-Kalman Filter	70%
					• Optical Flow	73%
					• Aggregate Channel Features (ACF)	95%

In comparison to the other algorithms, the results show that employing Fast R-CNN as an algorithm with SSD and YOLO techniques produces the greatest results in vehicle detecting systems.

*Table 6-d. Findings of **pedestrian detection** system*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
1	(Qu et al., 2018)	2018	• Color image conversion • Retinex image enhancement	MAE	• YOLOv3	94%
2	(Zahid et al., 2019)	2019	• Camera • Region-Based Convolutional Neural Network (RCNN) • SSD • YOLO	Confusion Matrix and IoU	• CNN • MobileNets	99.99%
3	(Kim et al., 2020)	2020	• Optical Camera • HOG • Hybrid Metaheuristic Pedestrian Detection (HMPD)	Confusion Matrix and F1-score	• SVM • CNN	98.5%
4	(Preethaa and Sabari, 2020)	2020	• HOG • Feature Vector of Image	Confusion Matrix and F1-score	• SVM-Based Enhanced Feature Extraction (SVMEFE) • HMPD	98.85%
5	(Pranav and Manikandan, 2020)	2020	• Camera (Real-time video captured)	Not Available	• CNN	• Min: 96.73% • Max: 100%

The pedestrian detection system can get improved results by combining CNN and MobileNets algorithms with the approaches R-CNN, SSD, and YOLO, as shown in Table 6-d.

*Table 6-e. Results of **lane detection and departure** system*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
1	(Zang et al., 2018)	2018	• Random Sample Consensus (RANSAC) • Kalman Filters • Gaussian Sum Particle Filter (GSPF) • Geometric Overture for Lane Detection by Intersections Entirety (GOLDIE)	Confusion Matrix	• Fully CNN	97.5%

*Table 6-e. Results of **lane detection and departure** system (continued)*

No.	Paper	Year	Technique(s)	Performance Metric(s)	Algorithm(s)	Overall Accuracy
2	(Liu et al., 2018)	2018	• Fuzzy Linear Discriminant Analysis (LDA) • Hough Transform • Inverse Perspective Mapping (IPM)	IoU	• Combining IPM and K-means Clustering	Above 90%
3	(Pizzati et al., 2019)	2019	• LiDAR • Stereo Camera • Instance Segmentation	mIoU (mean Intersection over Union)	• Cascaded CNN • Mask R-CNN	95.24%
4	(Olanrewaju et al., 2019)	2019	• HoG • Polynomial Regression • Hybrid Regression and Semantic Segmentation	MAE	• MobilenetV2	91.83%
5	(Aravind et al., 2020)	2020	• Optical Cameras • Fully Convolutional Network (FCN)	Not Available	• Dynamic Mode Decomposition (DMD)	98.03%

When compared to the other methods, the accuracy results show that using FCN methods with the DMD algorithm is the best.

5. Conclusions and Recommendations

In conclusion, seven major academic databases have been used (IEEE Xplore, Elsevier, Web of Knowledge, INSPEC, Springer, ACM Digital Library, ScienceDirect) to discover relevant literature that had recently been published. The total number of publications that have been studied exceeds 200, but only 70 of them were chosen according to the criteria of up-to-date information, the most significant AV system, applying vision approaches and IoT technologies, using smart sensors and actuators.

Many people are unclear about when autonomous vehicles will become effective in addressing transportation issues. According to optimistic predictions, autonomous cars may become sufficiently reliable, inexpensive, and commonplace by 2030–2035 to supplant most human driving, resulting in significant cost savings and advantages (Litman, 2020). However, there are valid reasons to be cautious. One of the main aspects of the motivation for the development of autonomous vehicles was to reduce the number of traffic accidents and to eliminate human factors as a cause of accidents. The idea is to have an autonomous vehicle that would be more dependable than a human. Vehicles must not only copy human behavior but also outperform humans to achieve this high-demand task.

The IoV is a subset of the IoT. It has become a critical platform for information exchange between vehicles, humans, and roadside infrastructure. The IoV is soon to become an essential part of our lives, allowing us to construct intelligent transportation systems free of road accidents, traffic lights, and other related issues. It would provide millions of people with more helpful, comfortable, and safe traffic services. With the rise of autonomous vehicles in recent years, a new trend has developed to use a variety of smart approaches and technologies to increase the performance and quality of autonomous decision-making. Combining various computer vision techniques with IoT in AV solutions, results in a high-performance embedded system that can be deployed in the environment to provide a more dynamic and resilient control system. Furthermore, computer vision has the potential to obtain higher-level information than other systems that rely on data from radar, lidar, ultrasonic, and other sensors. This is because these greater levels of information are important for designing systems in complicated environments, such as driver monitoring, pedestrian detection, traffic sign recognition, and so on. As has been discussed above, selecting the appropriate vision approach is entirely dependent on the type of AV system and the system's parameters. In the pedestrian detection system, for example, it was found that CNN and MobileNets are the best algorithms to use.

Compliance with Ethical Standards

Conflicts of Interest/ Competing Interests: The authors declare that the publishing of this review article does not involve any conflicts of interest.

References

Agarwal, P. K., Kumar, P., and Singh, H., 2020. Causes and Factors in Road Traffic Accidents at a Tertiary Care Center of Western Uttar Pradesh. In *Medico Legal Update*, 20(1): 38-4.

Ahmad, I., and Pothuganti, K., 2020. Design and Implementation of Real Time Autonomous Car by using Image Processing and IoT. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 107-113.

Aravind, H., Sivraj, P., and Ramachandran, K. I., 2020. Design and Optimization of CNN for Lane Detection. *In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 1-6.

Banerjee, A., Chakraborty, C., Kumar, A., and Biswas, D., 2020. Emerging Trends in IoT and Big Data Analytics for Biomedical and Health Care Technologies. In *Handbook of Data Science Approaches for Biomedical Engineering*, 121-152.

Cao, J., Song, C., Peng, S., Xiao, F., and Song, S., 2019. Improved Traffic Sign Detection and Recognition Algorithm for Intelligent Vehicles. *in Sensors*, 19(18): 4021-4042.

Celesti, A., Galletta, A., Carnevale, L., Fazio, M., Ĺay-Ekuakille, A., and Villari, M., 2018. An IoT Cloud System for Traffic Monitoring and Vehicular Accidents Prevention Based on Mobile Sensor Data Processing. *in IEEE Sensors Journal*, 18(12): 4795-4802.

Cheng, X., Zhang, R., and Yang, L., 2019. Wireless Toward the Era of Intelligent Vehicles. In *IEEE Internet of Things Journal*, 6(1): 188-202.

Chetouane, A., Mabrouk, S., Jemili, I., and Mosbah, M., 2020. Vision-based Vehicle Detection for Road Traffic Congestion Classification. *Concurrency and Computation: Practice and Experience*, e5983.

Choubey, P. C. S., and Verma, R., 2020. Vehicle Accident Detection, Prevention and Tracking System. *International Research Journal of Engineering and Technology (IRJET)*, 7(8): 2658-2663.

Ciuntu, V., and Ferdowsi, H., 2020. Real-Time Traffic Sign Detection and Classification Using Machine Learning and Optical Character Recognition. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, 480-486.

Da Xu, L., He, W., and Li, S., 2014. Internet of Things in Industries: A Survey, *IEEE Transactions on Industrial Informatics*, 10(4): 2233-2243.

Dhawale, R. Y., and Gavankar, N. L., 2019. Lane Detection and Lane Departure Warning System using Color Detection Sensor. *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Kannur, Kerala, India.

Eddie, E. E., Fabricio, D. E., Franklin, M. S., Paola, M. V., and Eddie D. G., 2018. Real-time Driver Drowsiness Detection based on Driver's Face Image behavior using a System of Human Computer Interaction Implemented in a Smartphone. In *International Conference on Information Technology & Systems*. Springer, Cham.

Elbasani, E., Siriporn, P., and Choi, J. S., 2020. A Survey on RFID in Industry 4.0. In *Internet of Things for Industry 4.0*, Springer, Cham, 1-16.

Elbery, A., Hassanein, H. S., Zorba, N., and Rakha, H. A., 2020. IoT-Based Crowd Management Framework for Departure Control and Navigation. In *IEEE Transactions on Vehicular Technology*.

Fernando, W. U. K., Samarakkody, R. M., and Halgamuge, M. N., 2020. Smart Transportation Tracking Systems based on the Internet of Things Vision. In *Connected Vehicles in the Internet of Things*, Springer, Cham, 143-166.

Garg, T., Kagalwalla, N., Churi, P., Pawar, A., and Deshmukh, S., 2020. A Survey on Security and Privacy Issues in IoV. *International Journal of Electrical & Computer Engineering*, l0(5): 5409, 5419.

Hamid, K., Bahman, J., Mahdi, J. M., and Reza, N. J., 2020. Artificial Intelligence and Internet of Things for Autonomous Vehicles. In *Nonlinear Approaches in Engineering Applications*, Springer, Cham, 39-68.

Hanan, E., 2019. Internet of Things (IoT), Mobile Cloud, Cloudlet, Mobile IoT, IoT Cloud, Fog, Mobile Edge, and Edge Emerging Computing Paradigms: Disambiguation and Research Directions. *Journal of Network and Computer Applications*, 128: 105-140.

Hirz, M., and Walzel, B., 2018. Sensor and Object Recognition Technologies for Self-Driving Cars. *in Computer-Aided Design and Applications*, 15(4): 501-508.

Hsu, S., Huang, C. L., and Cheng, C. H., 2018. Vehicle Detection using Simplified Fast R-CNN. In *2018 International Workshop on Advanced Image Technology (IWAIT)*, IEEE, 1-3.

Izquierdo-Reyes, J., Ramirez-Mendoza, R. A., Bustamante-Bello, M. R., Navarro-Tuch, S., and Avila-Vazquez, R., 2018. Advanced Driver Monitoring for Assistance System (ADMAS). *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 12(1): 187-197.

Kang, J., Yu, R., Huang, X., and Zhang, Y., 2018. Privacy-Preserved Pseudonym Scheme for Fog Computing Supported Internet of Vehicles. In *IEEE Transactions on Intelligent Transportation Systems*, 19(8): 2627-2637.

Khalifa, A. B., Alouani, I., Mahjoub, M. A., and Amara, N. E. B., 2020. Pedestrian Detection Using a Moving Camera: A Novel Framework for Foreground Detection. In *Cognitive Systems Research*, 60: 77-96.

Kim, B., Yuvaraj, N., Sri, P., Santhosh, K. R. R., and Sabari, A., 2020. Enhanced Pedestrian Detection using Optimized Deep Convolution Neural Network for Smart Building Surveillance. *Soft Computing*, 24: 17081-17092.

Kocić, J., Jovičić, N., and Drndarević, V., 2018. Sensors and Sensor Fusion in Autonomous Vehicles. *2018 26th Telecommunications Forum (TELFOR)*, Belgrade, 420-425.

Koresh, M. H. J. D., and Deva, J., 2019. Computer Vision Based Traffic Sign Sensing for Smart Transport. *Journal of Innovative Image Processing (JIIP)*, 1(1): 11-19.

Kumar, A., and Patra, R., 2018. Driver drowsiness monitoring system using visual behaviour and machine learning. *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, 339-344.

Lex, F., Daniel, E. B., Michael, G., William, A., Spencer, D., Bendikt, J., Jack, T., Aleksandr, P., Julia, K., Li, D., Sean, S., Alea, M., Andrew, S., Anthony, P., Bobbie, D. S., Linda, A., Bruce, M., and Bryan, R., 2019. MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction with Automation. In *IEEE Access*, 7: 102021-102038.

Li, G., Wang, Z., Fei, X., Li, J., Zheng, Y., Li, B., and Zhang, T., 2021. Identification and Elimination of Cancer Cells by Folate-Conjugated CdTe/CdS Quantum Dots Chiral Nano-Sensors. *Biochemical and Biophysical Research Communications*, 560: 199-204.

Li, W. L., Li, X. G., Qin, Y. Y., Ma, D., Cui, W., and Province, J., 2020. Real-time Traffic Sign Detection Algorithm Based on Dynamic Threshold Segmentation and SVM. *Journal of Computers*, 31(6): 258-273.

Litman, T., 2020. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning.

Liu, J., Lou, L., Huang, D., Zheng, Y., and Xia, W., 2018. Lane Detection based on Straight Line Model and K-means Clustering. In *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, IEEE, 527-532.

Liu, Z., Peng, Y., and Hu, W., 2020. Driver Fatigue Detection based on Deeply-Learned Facial Expression Representation. In *Journal of Visual Communication and Image Representation*, 71: 102723.

Lu, H., Liu, Q., Tian, D., Li, Y., Kim, H., and Serikawa, S., 2019. The Cognitive Internet of Vehicles for Autonomous Driving. In *IEEE Network*, 33(3): 65-73.

Minhas, A. A., Jabbar, S., Farhan, M., and ul Islam, M. N., 2019. Smart Methodology for Safe Life on Roads with Active Drivers based on Real-Time Risk and Behavioural Monitoring. *Journal of Ambient Intelligence and Humanized Computing*, Springer, 1-13.

Minovski, D., Åhlund, C., and Mitra, K., 2020. Modeling Quality of IoT Experience in Autonomous Vehicles. *IEEE Internet of Things Journal*, 7(5): 3833-3849.

Narote, S. P., Bhujbal, P. N., Narote, A. S., and Dhane, D. M., 2018. A Review of Recent Advances in Lane Detection and Departure Warning System. In *Pattern Recognition*, 73: 216-234.

Ng, J. R., Wong, J. S., Goh, V. T., Yap, W. J., Yap, T. T. V., and Ng, H., 2019. Identification of Road Surface Conditions using IoT Sensors and Machine Learning. *Computational Science and Technology*, Springer, Singapore, 481.

Olanrewaju, R. F., Fakhri, A. S. A., Sanni, M. L., and Ajala, M. T., 2019. Robust, Fast and Accurate Lane Departure Warning System using Deep Learning and Mobilenets. In *2019 7th International Conference on Mechatronics Engineering (ICOM)*, IEEE, 1-6.

Qu, H., Yuan, T., Sheng, Z., and Zhang, Y., 2018. A Pedestrian Detection Method based on YOLOv3 Model and Image Enhanced by Retinex. *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, IEEE.

Philip, J. M., Durga, S., and Esther, D., 2021. Deep Learning Application in IoT Health Care: A survey. In *Intelligence in Big Data Technologies-Beyond the Hype*, Springer, Singapore, 199-208.

Pizzati, F., Allodi, M., Barrera, A., and García, F., 2019. Lane Detection and Classification using Cascaded CNNs. In *International Conference on Computer Aided Systems Theory*, Springer, Cham, 95-103.

Pranav, K. B., and Manikandan, J., 2020. Design and Evaluation of a Real-time Pedestrian Detection System for Autonomous Vehicles. In *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*, IEEE, 155-159.

Preethaa, K. S., and A. Sabari, A., 2020. Intelligent Video Analysis for Enhanced Pedestrian Detection by Hybrid Metaheuristic Approach. *Soft Computing*, 1-9.

Rahul, P. K., Abhishek, S. K., Zeeshaan, W. S., Vaibhav, U. B., and Nasiruddin, M., 2020. IoT based Self Driving Car. *International Research Journal of Engineering and Technology (IRJET)*, 7(3): 5177-5181.

Rani, R., Kumar, N., Khurana, M., Kumar, A., and Barnawi, A., 2021. Storage as a Service in Fog Computing: A Systematic Review. *Journal of Systems Architecture*, 102033.

Ravikumar, S., and Kavitha, D., 2021. IOT based Autonomous Car Driver Scheme based on ANFIS and Black Widow Optimization. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.

Sadiku, M. N., Tembely, M., and Musa, S. M., 2018. Internet of Vehicles: An Introduction. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(1): 11.

Sakhare, K. V., Tewari, T., and Vyas, V., 2020. Review of Vehicle Detection Systems in Advanced Driver Assistant Systems. *Archives of Computational Methods in Engineering*, 591-610.

Sastry, A., Bhargav, K., Pavan, K. S., and Narendra, M., 2020. Smart Street Light System using IoT. *International Research Journal of Engineering and Technology (IRJET)*, 7(3): 3093-3097.

Savaş, B. K., and Becerikli, Y., 2020. Real Time Driver Fatigue Detection System based on Multi-Task ConNN. In *IEEE Access*, 8: 12491-12498.

Sharma, S., 2021. Towards Artificial Intelligence Assisted Software Defined Networking for Internet of Vehicles. In *Intelligent Technologies for Internet of Vehicles*. Springer, Cham, 191-222.

Shashank, P., Pavan, M., Niharika, P., Jalakam, S., and Sathweek, B., 2020. Machine Learning Based Water Quality Checker and pH Verifying Model. *European Journal of Molecular & Clinical Medicine*, 7(11): 2220-2228.

Singh, D., Tripathi, G., and Jara, A.J., 2014. A Survey of Internet-of-Things: Future Vision, Architecture, Challenges and Services, *2014 IEEE World Forum on Internet of Things (WF-IoT).* IEEE.

Srivastava, M., and Kumar, R., 2021. Smart Environmental Monitoring Based on IoT: Architecture, Issues, and Challenges. In *Advances in Computational Intelligence and Communication Technology*, Springer, 349-358.

Sultana, T., and Wahid, K. A., 2019. IoT-Guard: Event-Driven Fog-Based Video Surveillance System for Real-Time Security Management. In *IEEE Access*, 7: 134881-134894.

Tsai, C., Ching-Kan, T., Ho-Chia, T., and Jiun-In, G., 2018. Vehicle Detection and Classification based on Deep Neural Network for Intelligent Transportation Applications. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 1605-1608.

Uma, S., and Eswari, R., 2021. Accident Prevention and Safety Assistance using IOT and Machine Learning. *Journal of Reliable Intelligent Environments*, 1-25.

Varma, B., Sam, S., and Shine, L., 2019. Vision Based Advanced Driver Assistance System Using Deep Learning. 2019 *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 1-5.

Wang, J., Lim, M. K., Wang, C., and Tseng, M. L., 2021. The evolution of the Internet of Things (IoT) over the past 20 years. *Computers & Industrial Engineering*, 155: 107174.

Wei, Y., Tian, Q., Guo, J., Huang, W., and Cao, J., 2019. Multi-Vehicle Detection Algorithm through Combining Harr and HOG Features. *Mathematics and Computers in Simulation*, 155: 130-145.

Whitmore, A., Agarwal, A., and Da Xu, L., 2015. The Internet of Things-A Survey of Topics and Trends, *Information Systems Frontiers*, 17(2): 261-274.

World Health Organization (WHO)., 2020. Global Status Report on Road Safety. Accessed on: 28th October 2021. [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/report/en/.

Xiang, X., Zhai, M., Lv, N., and El Saddik, A., 2018. Vehicle Counting based on Vehicle Detection and Tracking from Aerial Videos. In *Sensors*, 18(8): 2560.

Zahid, A., Iniyavan, R., and Madhan, M. P., 2019. Enhanced vulnerable pedestrian detection using deep learning. *International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 0971-0974.

Zaidan, A. A., and Zaidan, B. B., 2020. A Review on Intelligent Process for Smart Home Applications based on IoT: Coherent Taxonomy, Motivation, Open Challenges, and Recommendations. *Artificial Intelligence Review*, 53(1): 141-165.

Zang, J., Zhou, W., Zhang, G., and Duan, Z., 2018. Traffic Lane Detection using Fully Convolutional Neural Network. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 305-311.

Zhang, J., Xie, Z., Sun, J., Zou, X., and Wang, J., 2020. A Cascaded R-CNN with Multiscale Attention and Imbalanced Samples for Traffic Sign Detection. In *IEEE Access*, 8: 29742-29754.

Zhao, C., Li, L., Xin, P., Zhiheng, L., Fei-Yue, W., and Xiangbin, W., 2021. A Comparative Study of State-of-the-art Driving Strategies for Autonomous Vehicles. *Accident Analysis & Prevention*. 150: 105937.

Authors's Biography

	Mohammed Qader Kheder was born in Ranyah/ Sulaimani city- Kurdistan Region/ Iraq. He earned a BSc degree at the University of Sulaimani (2008), MSc degree in the United Kingdom in advanced computer science (2011) at the University of Huddersfield. He also by his researches could pass from Assistant Lecturer to Lecturer at the University of Sulaimani. He is a Ph.D. student whose study focuses on utilizing IoT technology and deep learning techniques in autonomous vehicles. His research interest includes IoT, machine learning, wireless access networks, and vision tactics.
	Prof. Aree Ali Mohammed was born in Sulaimani city–Kurdistan Region Iraq. He obtained a BSc degree at the University of Mousle (1995), an MSc degree in France in Computer Science (2003), and a Ph.D. In multimedia systems at the Univesity of Sulaimani (2008). He directed Information Technology Directorate for four years (2010– 2014) and the head of the computer science department/ college of Science / University of Sulaimani for seven years. The main field of interest is multimedia system applications for processing, compression, and security. Many papers have been published in scientific journals throughout the world.

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
eISSN: 2255-2863 - DOI: https://doi.org/10.14201/ADCAIJ2022113 - CDU: 004 -
IBIC: Computación e informática (U) - BIC: Computing & Information Technology (U) -
BISAC: Computers / General (COM000000)

Regular Issue, Vol. 11, N. 3 (2022)

INDEX

Ediciones Universidad Salamanca

RESEARCH GROUP BISITE .usal.es