Health Care Application using Machine Learning and Deep Learning

Ajinkya Padule[1], Aman Patel[2], Arsalan Patel[3], Aman Shaikh[4], Jyoti Gavhane[5]

Department of Computer Science and Engineering, MIT School of Engineering,
MIT Art, Design and Technology University, Pune - 412201, India

---***---

Abstract - *Machine learning techniques are widely used in a variety of fields, and the health care industry, in particular, has benefited greatly from machine learning prediction methods. Because disease prediction is a difficult task, it is necessary to automate the process in order to avoid the risks associated with it and to alert the patient ahead of time. Doctors require precise predictions of their patients' disease outcomes. Furthermore, timing is a significant factor that influences treatment decisions for accurate predictions. Early disease prediction has become an important task. However, doctors find it difficult to make accurate predictions based on symptoms. The most difficult task is correctly predicting disease. Machine learning is being used more and more in the field of medical diagnosis. This can be attributed primarily to advancements in disease classification and recognition systems, which can provide data that aids medical experts in the early detection of fatal diseases, thereby significantly increasing patient survival rates. In this paper, we present an intelligent healthcare application that can predict and provide information on various diseases.*

Key Words: Machine Learning, Deep Learning, AI, Healthcare, Disease Prediction

1. INTRODUCTION

Health care is widely recognized as an important determinant in promoting people's overall physical, mental, and social well-being, and it can add significantly to a nation's economy, development, and industrialization. Some people don't get the necessary health care they require because they do not have health insurance or because they live too far away from providers who provide them. More people can get the care they need if initiatives to increase access to health care services are implemented, such as lowering costs, increasing the use of telehealth, and improving insurance coverage.

Healthcare is one of the fastest-growing industries in today's economy; more people need care, and it's getting more expensive. Government spending on healthcare has reached an all-time high, despite the obvious need for improved patient-physician interaction. Big data and machine learning technologies have the potential to benefit both patients and providers in terms of better care and lower costs [1]. Doctors believe that time is a key element in diagnosis, and arriving at a suitable conclusion in a timely manner may significantly benefit patients. As a result, accurate patient outcome prediction is an issue in healthcare.

Artificial Intelligence (AI) has recently become popular in medicine and the healthcare industry. AI has made computers smarter and capable of thinking. Machine learning is considered a subfield of AI in numerous research studies.

1.1 Machine Learning (ML)

Machine learning techniques are widely applied in various fields, and the health care sector, in particular, has benefited greatly from machine learning prediction techniques. Its goals include correctly predicting diseases, improving medical treatment, and improving clinical outcomes. In medical applications, machine learning algorithms can help doctors make better treatment decisions for patients by utilizing an effective healthcare system. This may be attributed mostly to advancements in disease classification and identification systems, which can give data that supports medical specialists in the early discovery of lethal diseases, resulting in a considerable rise in patient survival rates.

Machine Learning: the classic definition is - A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [2].

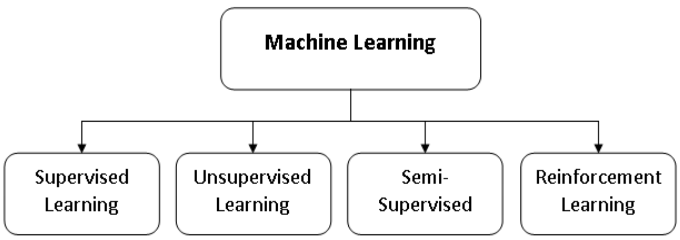

Fig -1: Machine Learning Types

Machine Learning Algorithms:

- Logistic Regression

Logistic Regression is a classification process for determining the probability of an event occurring or not. It is used to represent a binary or categorical outcome that has only two classes. It is identical to linear regression, with the exception that the variable's result is categorical rather than continuous. For prediction, it uses the Logit Link function, which fits the data values.

- Decision Tree

The Decision tree is a machine learning approach that is used for both classification and regression analysis. It is identical to the tree analogy in real life. It's a tree-like graph that starts with a single node and branches out to all of the outcomes. A decision tree, unlike linear models, is supervised learning that maps non-linear relationships as well.

- Random Forest

A random forest is a collection of multiple decision trees. The random forest classifier is essentially a collection of decision tree classifiers, each of which is built using several random vectors and can vote for the most favored prediction class. The addition of randomization to the model prevents it from overfitting, resulting in improved classification results.

- Support Vector Machines (SVM)

Support Vector Machines, also known as Support Vector Networks, are supervised learning algorithms that can be used for regression and classification problems. It uses parallel lines called the hyperplane to divide data points plotted in a multidimensional space into categories. The maximization of the margin between the hyperplane is required for data point classification.

1.2 Deep Learning (DL)

Deep learning is a type of machine learning and artificial intelligence (AI) that closely resembles how humans acquire specific types of knowledge. It uses artificial neural networks, which are designed to imitate how humans think and learn, as compared to machine learning, which uses simpler concepts. A weighted and bias-corrected input value is passed through a non-linear activation function such as ReLu and softmax to generate an output in a traditional Deep Neural Network (DNN).[20] As a result, the goal of training a DNN is to optimize the network's weights in order to minimize the loss function. [3]

Convolutional Neural Network (CNN):

A Convolutional Neural Network, or CNN, is a type of artificial neural network used for image/object recognition and classification in Deep Learning. Using CNN, Deep Learning identifies objects in an image. A CNN is comprised of three major layers: convolution layer, pooling layer, and fully connected layer.

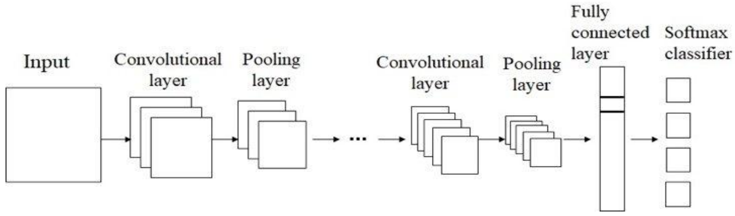

Fig -2: CNN Architecture

Transfer Learning:

- Transfer learning is a process in which a model trained on one problem is used in some way on another related problem.
- Transfer learning is a deep learning technique in which a neural network model is first trained on a problem similar to the one being solved.
- Transfer learning reduces the training time for a neural network model and can lead to lower generalization error.

VGG16:

Convolution layers of 3x3 filters with stride 1 and always used the same padding and max pool layer of 2x2 filters with stride 2. The 16 in VGG16 refers to it having 16 weighted layers. This network is quite large, with approximately 138 million (approx) parameters.

VGG16 is a convolutional neural network architecture which stands for Visual Geometry Group and it is also known as OxfordNet. It was proposed in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" by Karen Simonyan and Andrew Zisserman from the University of Oxford.

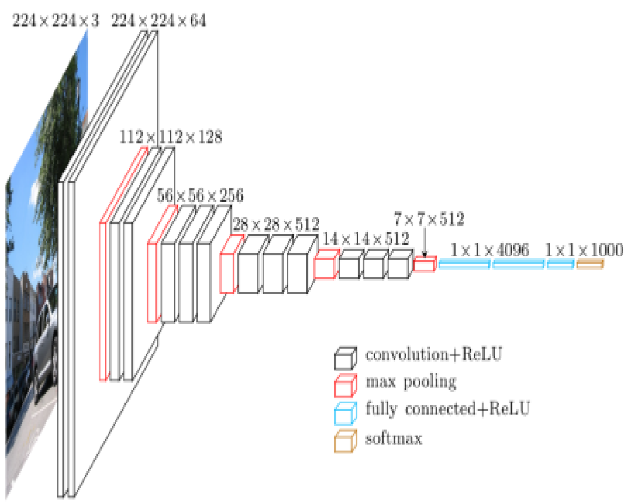

Fig -3: VGG16 Architecture

2. Literature Survey

Apurb Rajdhan [4] and others in their paper applied various ML algos for prediction of heart disease using UCI machine learning repository dataset. Algos used were Decision Tree, Logistic Regression, Random Forest and Naive Bayes. Out of which RF gave highest accuracy of 90.16%.

The authors [5] in their paper have build a web based application for heart disease prediction. They used UCI dataset obtained from University of California, Irvine. They used SVM, LR and NB algos and SVM was more accurate than other two with 64.4% accuracy.

In [6], Mujumdar and other authors, they used various ML Algos for diabetes prediction. LR gave highest accuracy of 96%.

In this, authors have used PIMA Indians diabetes dataset from UCI. They analyzed both ML and DL Algos for prediction of diabetess. The results of RF was more efficient which gave accuracy of 83.67%. [7]

The authors of [8] used dataset from UCI ML Reposiory.In addition original dataset was collected from northeast of Andhra Pradesh, India. For liver disease prediction, they used 6 ML algos. And in terms of accuracy, LR achieved highest accuracy of 75%.

In this, they used Naïve Baiyes and SVM algorithms for liver disease prediction. From the experimental results, the SVM classifier is considered as a best algorithm because of its highest classification accuracy. On the other hand, while comparing the execution time, the Naïve Bayes classifier needs minimum execution time. Dataset used was Indian Liver Patient Dataset (ILPD).[9]

In this paper, a deep learning algorithm was developed to accurately predict Malaria in a short amount of time. Three CNN models were built, with the highest accuracy model being chosen. In comparison to other CNN models, the Fine-Tuned CNN had a high accuracy rate.[10]

The data set used was obtained from the US National Library of Medicine in the and contains 27,558 cell pictures. They used the CNN technique and achieved 95% accuracy.[11]

In [12], the authors have predicted Pneumonia disease and used the dataset provided by Guangzhou Women and Children's Medical Center Guangzhou which is openly available on Kaggle and has 5856 images of chest X-ray. They used grayscale images of size 200*200 pixels. Data augumentation was performed on the dataset for balancing the dataset. Achieved accuracy of 88.90%.

3. Proposed System

This section explains the methods, algorithms, and system architecture that we used to develop our application. We build a web-based application that predicts the disease In this study, we have used ML and DL algorithms to create a healthcare system to predict different diseases. The diseases for which we proposed our system are Diabetes, Heart, Liver, Malaria, and Pneumonia. Out of which, for Diabetes, Heart, and Liver, we used ML and for Malaria and Pneumonia, we used DL. The proposed work predicts diseases by exploring the ML algorithms and doing performance analysis. The objective of this study is to effectively predict if the patient suffers from the disease. The input values from the patient's health report are entered by the user. The data is fed into the ML model which predicts the probability of having the disease.

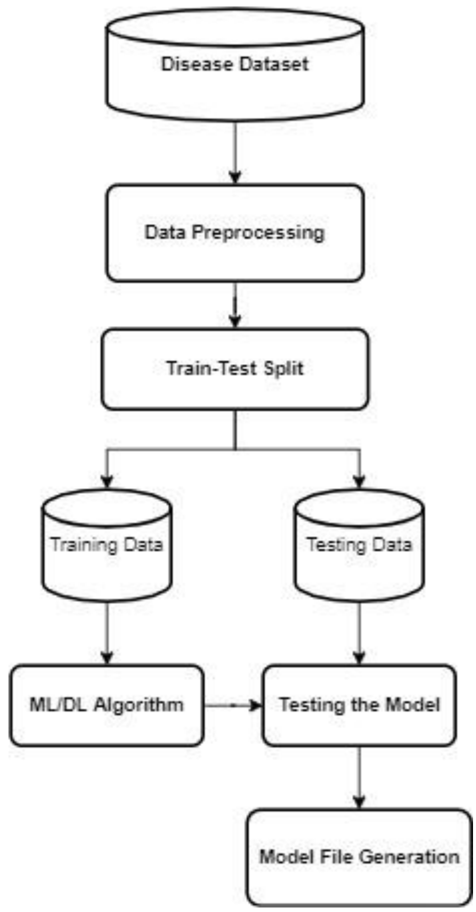

Fig - 4: Model File Generation

We have divided the model training into ML training and DL training for better understanding.

ML Training:

Datasets for Diabetes, Heart and Liver diseases were collected from UCI Machine Learning Reposiory in Comma-Seperated Values (CSV) format. After collecting data, we performed some data preprocessing on that data like cleaning data, normalizing, standardizing, handling null values. Then we split the dataset into training and testing. 80% of the data was used for training the model and the remaining 20% was used for testing the model.

Training dataset is the dataset which is used to train a model. Testing dataset is used to check the performance of the trained model.

For training the model files, we used various ML algos. We used Logistic Regression, Random Forest, SVM, Decision Tree. All these algos were applied to every disease dataset and we evaluated their accuracies. For each of the algorithms the performance is computed and analyzed.

DL Training:

For Malaria disease, we have used Malaria Cell Images Dataset dataset from Kaggle and for Pneumonia, we have used Chest X-Ray Images (Pneumonia) dataset from Kaggle. We applied some preprocessing on the images such as rescale, zoom range, shear range, horizontal flip, etc. Image size was 224*224. We used 27581 training data and 525 testing data for malaria disease and 5216 training data and 624 testing data for pneumonia disease. We have applied transfer learning here. VGG16 with weights initialized with imagenet was used. Then we trained our model for 25 epochs.

System Architecture:

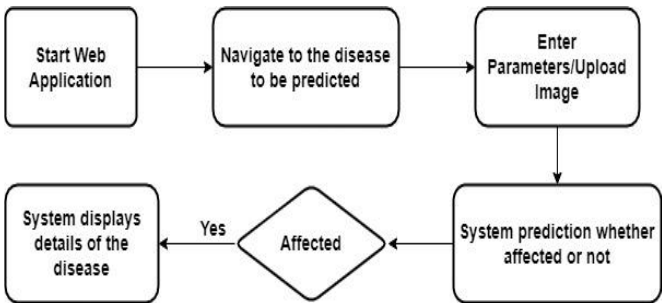

Fig - 5: System Architecture

Front-End:

Front-End is what we call User-Interface. It represents how the user interacts with the system. For our system, we have used HTML-CSS-Javascript. HTML(HyperText Markup Language) is said to be the skeleton, CSS(Cascading Style Sheets) is considered the body, and Javascript is the brain of the design system.

The user is required to input the details of the patient data through the front-end which passes this data to the backend where the prediction takes place. For diabetes, heart, and liver, the frontend provides the input field to enter the values and for malaria and pneumonia, the frontend allows the user to upload the image. Then the results from the backend are displayed on the frontend. And if the person is affected, then the system displays the details of that disease which will help in diagnosis and understanding of the severity of the disease.

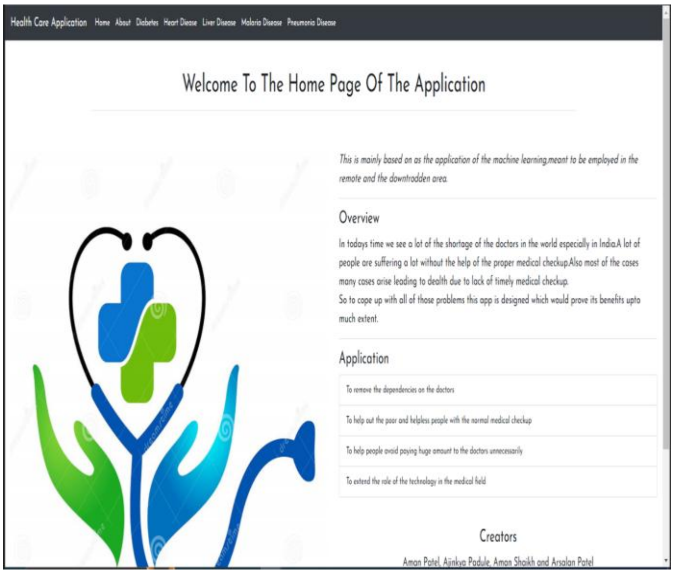

Fig - 6: Output 1-Home Screen

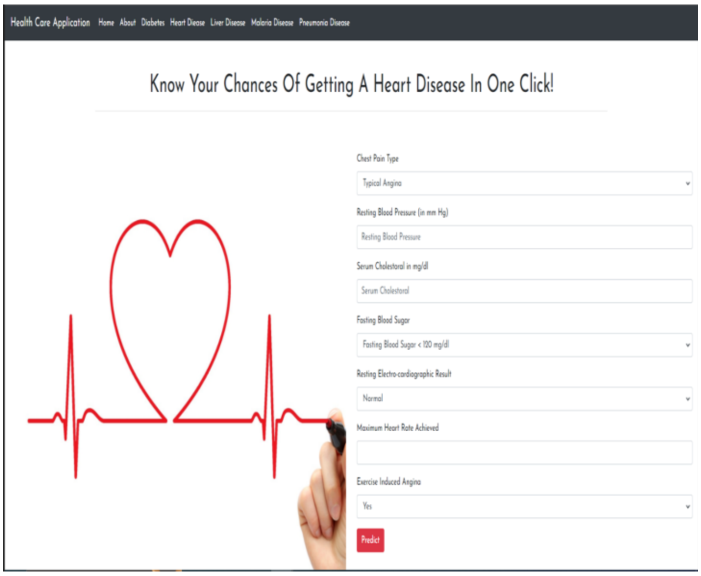

Fig - 7: Output 2- Input Patient Details

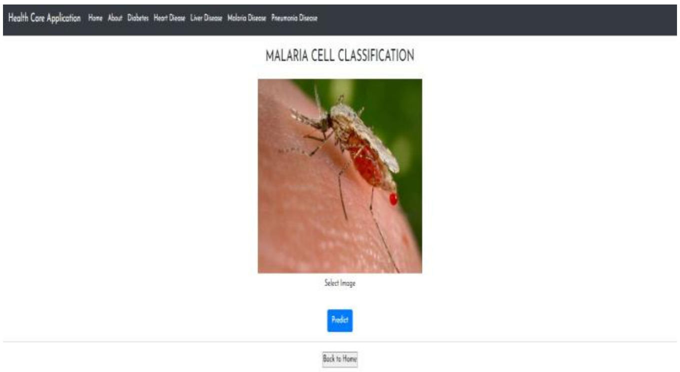

Fig - 8: Output 3-Input Patient Image

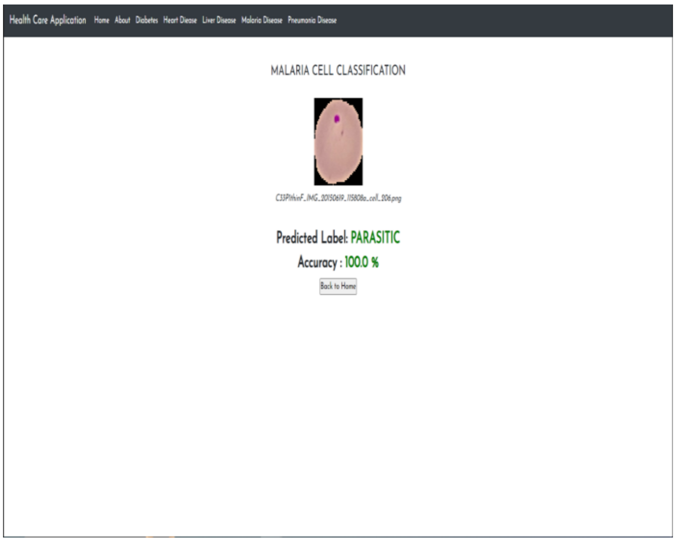

Fig - 9: Output 4- Disease Prediction Output Screen

Back-end:

The backend is where the logic gets executed. It operates in the background of the system. The framework which we have used to develop our application is flask which is programmed using python. Python is also used for the model files generation.

For the deployment of our application, we have used Heroku.

4. Results:

This section will discuss the outcomes of the proposed system.

The results obtained by applying Random Forest, Decision Tree, SVM and Logistic Regression are shown in this section.

Table -1: Diabetes, Heart and Liver Diseases Accuracies

Disease	Algorithm	Accuracy
Diabetes	Logistic Regression	81.01
	Random Forest	**84.81**
	SVM	81
	Decision Tree	75.95
Heart	***Logistic Regression***	**80.26**
	Random Forest	72.36
	SVM	81.57
	Decision Tree	61.84
Liver	Logistic Regression	65.5
	Random Forest	**83.33**
	SVM	66.56
	Decision Tree	60.2

After analyzing the results obtained from the algorithms, we inferred that for diabetes disease random forest classifier gave the best accuracy, for heart disease SVM performed well and gave the highest accuracy and for liver disease random forest classifier gave better accuracy.

Table -2: Best Disease Accuracies Selected

Disease	Algorithm	Accuracy
Diabetes	Random Forest Classifier	84.01
Heart	SVM	81.57
Liver	Random Forest	83.33
Malaria	VGG16	94.29
Pneumonia	VGG16	95.48

5. CONCLUSIONS

With the rising number of deaths caused by various diseases, it has become necessary to design a system that can efficiently and reliably anticipate diseases. The study's goal was to discover the best effective ML and DL algorithms for detecting these illnesses. Using the UCI machine learning repository dataset, this study analyses the accuracy scores of Decision Tree, Logistic Regression, Random Forest, SVM, VGG16 algorithms for predicting different disease. In this system application, we successfully predicted 5 diseases that are Diabetes, Heart, Liver, Malaria and Pneumonia. By providing the input as a patient record, we were able to get an accurate general illness risk prediction as an output, which let us understand the degree of disease risk prediction.

The healthcare business is experiencing greater difficulties and is growing more costly. To address these difficulties, a variety of machine learning and deep learning methods are implemented. As a result, such integration should be encouraged for the sake of humanity's advancement.

REFERENCES

[1] R. Bhardwaj, A. R. Nambiar and D. Dutta, "A Study of Machine Learning in Healthcare," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 2017, pp. 236-241, doi: 10.1109/COMPSAC.2017.164.

[2] Machine Learning, Tom Mitchell, McGraw Hill, 1997.

[3] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, et al., Federated learning in mobile edge networks: a comprehensive survey, IEEE Commun. Surv. Tutorials (2020).

[4] Rajdhan, Apurb & Agarwal, Avi & Sai, Milan & Ghuli, Poonam. (2020). Heart Disease Prediction using Machine Learning. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS040614.

[5] Jagtap, A., Malewadkar, P., Baswat, O. and Rambade, H., 2019. Heart disease prediction using machine learning. International Journal of Research in Engineering, Science and Management, 2(2), pp.352-355.

[6] Mujumdar, Aishwarya & Vaidehi, V.. (2019). Diabetes Prediction using Machine Learning Algorithms. Procedia Computer Science. 165. 292-299. 10.1016/j.procs.2020.01.047.

[7] Yahyaoui, Amani & Rasheed, Jawad & Jamil, Akhtar & Yesiltepe, Mirsat. (2019). A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques. 10.1109/UBMYK48245.2019.8965556.

[8] Rahman, A. K. M. & Shamrat, F.M. & Tasnim, Zarrin & Roy, Joy & Hossain, Syed. (2019). A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms. 8. 419- 422.

[9] Mohan, Vijayarani. (2015). Liver Disease Prediction using SVM and Naïve Bayes Algorithms

[10] Shekar, Gautham; Revathy, S.; Goud, Ediga Karthick (2020). [IEEE 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) - Tirunelveli, India (2020.6.15-2020.6.17)] 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) - Malaria Detection using Deep Learning. , (), 746– 750. doi:10.1109/ICOEI48184.2020.9143023

[11] Sriporn, Krit & Tsai, Cheng-Fa & Tsai, Chia-En & Wang, Paohsi. (2020). Analyzing Malaria Disease Using Effective Deep Learning Approach. Diagnostics (Basel, Switzerland). 10. 10.3390/diagnostics10100744

[12] L. Račić, T. Popović, S. čakić and S. Šandi, "Pneumonia Detection Using Deep Learning Based on Convolutional Neural Network," 2021 25th International Conference on Information Technology (IT), 2021, pp. 1-4, doi: 10.1109/IT51528.2021.9390137.

[13] Ajinkya Padule , Aman Patel , Aman Shaikh , Arsalan Patel, Jyoti Gavhane, 2022, A Comparative Study on Health Care using Machine Learning and Deep Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 11, Issue 03 (March 2022).

[14] Aman Shaikh , Nilima Kulkarni , Aman Patel , Arsalan Patel, 2021, Intelligent Advertisement System, International Journal of innovative Research in Technology (IJIRT) Volume 7, Issue 8 (January 2021).