A HEALTHCARE PORTAL USING MACHINELEARNING

Manish Bhamare[1], Sayali Joshi[1], Shubham Sontakke[1*], Shubham Raut[1], Prof. Minal Zope[1]

Abstract –
Irregularities were seen in hospital prices during Covid-19 which caused a major problem for patients Many families lost membersdespite spending enormous quantities of money, and in some cases, hospitals required payment of unpaid bills before returning bodies. There is a need for an information portal that can regulate all private hospitals, provide accurate information about costs and medical facilities offered in private hospitals and direct citizens towards affordable medical emergency services in accordancewith their preferences and availability. Using Machine Learning we are integrating two more functionalities: a) Medical Prescription Reader: Often the prescriptions given by doctors are only readable to particular pharmacists. It can be misinterpreted due to the messy handwriting. We take an image as an input and preprocess it, then process it and classify an extracting feature byCNN and OCR in post-processing is applied. b) Disease Prediction: Many times, we visit a hospital for one reason, and it turns outto be something else and must visit another doctor or a specialist for treatment of a specific disease. We are providing a disease prediction tool which will predict the possibility of a certain disease so that users can visit a doctor specialized in that domain. Thisis possible with the help of machine algorithms such as Random Forest Naïve Bayes and SVM for predicting disease.

Keywords – Convolutional Neural Network (CNN), Random Forest, OCR, SVM, Naïve Bayes

[1]Computer Engineering, AISSMS Institute ofInformation Technology, Pune, India

[*]**Corresponding Author:** Shubham Sontakke
[*]Computer Engineering, AISSMS Institute ofInformation Technology, Pune, India,
shubhamsontakke110@gmail.com

DOI: 10.53555/ecb/2023.12.3.248

INTRODUCTION

According to history, pandemics occur once every century. But, when it comes, it usually comes with a majordisaster, a major strain on the economy, and a major strainon the healthcare system. In these scenarios, our healthcaresystem has a significant impact on the people of the country. The recent COVID-19 outbreak in India has had asignificant impact, and the early 2020 events have so vividly recalled the numerous pandemics that originated in or affected the Subcontinent over the previous two centuries—cholera in 1817, bubonic plague in 1896, influenza in 1918—that it has become commonplace, if notdownright facile, to assert that history is repeating itself. India was destined to suffer the most due to its largepopulation, high population densities, widespread hunger and poverty, millions of migrant laborers, and a failing public health infrastructure. The 1918 model would result in the death of millions of Indians. The COVID-19 epidemic has had a significant impact on people's physicaland mental well-being as well as the global economy. Despite having milder symptoms, lower morbidity, and a better prognosis than adults, young people are less prone tosevere forms of the illness. However, they have also seen an increase in stress, which has led to loneliness, anxiety, and depression in many. The impact of COVID-19 on young people's mental health has been researched since emotional symptoms in adolescence have been connected to a variety of major mental health outcomes, including suicide, long-term physical health repercussions, and asignificant healthcare burden. Data mining in the healthcare and medical sectors is more crucial than ever nowadays. Big databases can yield valuable information when particular data mining techniques are effectively applied, allowing medical practitioners to make speedier choices and improve patient care. The purpose of employing categorization and algorithms is to assist the clinician. The majority of people in today's society routinely misunderstand medical prescriptions, which leads to their taking the incorrect meds or the improper amount of pharmaceuticals, which is extremely dangerousto their health and in some cases could even be fatal. Patients and pharmacists frequently lack basic medical knowledge, making it impossible to interpret most doctors'handwriting, which adds to the issue. Despite the fact that it is a problem that arises more frequently, technology can solve it. Long-term effects could be more detrimental to health than the infection itself. Thus, identifying those whoare most susceptible to mental health troubles andmeasuring early indicators of mental health concerns in young people—such as fears and negative emotions—should be top priorities for academics and policymakers. Although both high- and low-income countries should consider the value of this information, early predictions of these demands may be especially helpful for developing nations. Countries have suffered considerably from inconceivable increases in the expense of medical services, in addition to mental health. A report that appeared in the Indian Express on April 15, 2021, stated, "The prices of Remdesivir were found to be exaggerated." Remdisivir was found to cost up to 30.000 rupees for each vial, compared toa government-approved price of 2000 rupees. We found inconsistencies in the delivery of private healthcare servicesin light of the previous article. In the case of price irregularities, this irregularity was found. When it came to bed availability, the same thing took place. Regarding the availability of beds, the particulars of the medical care provided, etc., there was inadequate information. Someone is unable to decide what is affordable when they unpredictably require emergency service. A national information portal is required to oversee all private hospitals, providing accurate data on prices and medical provided in private hospitals across the country, and assisting people in locating accessible emergency medical care based on their preferences and accessibility. This will make it harder for private hospitals totake advantage of the general people

LITERATURE SURVEY

Abeer T, Elsersy M, Ahmed Sherif, Hassan H, Omar Abdel salam, Khaled H. Almotairi have submitted a study that provides a review of several algorithms using machine learning and their uses in many fields, emphasizing the advantages of methods of ML for enhancing healthcare and building effective support infrastructure for the medical industries. The primary aim is to highlight the ML algorithms that have already been used in healthcare, to give the information needed to researchers willing to investigate MLin healthcare. It was difficult for healthcare practitioners to obtain and analyses the vast amounts of data for accurate forecasts and treatments in the past because there were no technology or tools accessible. Large dataset processing and categorization are now simple thanks to ML. This study shows how ML is important to our daily lives, particularly for healthcare industry. With the ML algorithm, we can predict and diagnose diseases early and establish the connection between drugs and diseases. (A. Thamara, 2021) Authors have argued in a report that because doctors are so busy today, they often write illegible

prescriptions, which makes it difficult to understand pharmaceutical names. Before buying their recommended medications, Patients might be curious to learn more about them, due to doctors' inconsistent handwriting and poor handwriting quality, techies are finding a solution to effectively deal the issue. We can now distinguish new handwriting thanks to ML, due to which the machine can learn several styles of handwriting for the same medication. With a smartphone app which can recognize written drugs, and deliver a medicine with dosage in digital text format, this research gave a solution which benefits both patients and pharmacists. The ML based System uses several pre-processing procedures to identify the names of the medications and their dosages for the gathered data set. This will assist in reducing instances of drug name distortion, supporting pharmacists in decreasing their uncertainties. (E. Hassan, 2021) Kushal Rashmikant Dalal has proposed implementation using ML in clinical settings,reflecting the exciting potential to enhance healthcare. When making decisions, adopting, and supporting medical areas, commercial organizations evaluate and leverage ML growth. The production of individualized medical experiences and the use of speech recognition are two instances where this powerful subdivision of AI is used. By examining the transformation, the sector because of this sector's capabilities, this paper seeks to critically analyze how it has been used to the healthcare sector. (K. R. Dalal, 2020) The idea of Medical Decision Support is emphasized to help physicians and serve as an additional validation tool throughout the diagnosis and illness prediction processes, according to Safiah Endargiri and Kaouther Laabidi. In this study, we contrast an automated diagnosis method forCOVID-19, and pneumonia along with healthy patients. The method incorporates inputs from CT scans and X-ray pictures. Support Vector Machine algorithm (SVM), K- Nearest Neighbor algorithm (KNN), discriminant, tree method and Naive Bayes algorithm are five different classification learners that are included in our flexible use of a hybrid method which combines two different pre-trained neural network types. (Laabidi, 2020) Dhiraj Dahiwade, Prof. Gajanan Patle, Prof. Ektaa Meshram statesthat people experience a multiple variety of illnesses whichresults in not only their lifestyle but also the environment. Therefore, predicting sickness sooner becomes a crucial duty. However, it is very challenging for doctors to make an exact prediction based just on patient symptoms. The most challenging task is accurately diagnosing a condition. To address this kind of problem, predictive analytics usingdata mining is essential. In order to accurately forecast diseases, we use the KNN and CNN Machine learning algorithms. (D. Dahiwade, 2019) Dr E.Kamalanaban, M. Gopinath, and S.Premkumar proposed in order to detect handwritten drug names and deliver intelligible digitallanguage, this article employs a smartphone application called Medicine Box to overcome issues with doctor's prescriptions. TensorFlow, a machine learning framework,and Custom Repository is used in this mobile application to make a match between the partial string and the name ofdrug. With Medicine Box, medicine name misunderstandings can be decreased. This helps pharmacists and makes it clear to regular people what the doctor has prescribed in the prescription. (Kamalanab anand E. Gopinathand, 2018) Sneha Grampurohit and Chetan Sagarnal proposed the creation and use of a numberof well-known data mining techniques in several practical application fields. Several areas, including the prediction of diseases, have made effective use of machine learning techniques. Making a classifier system with machine learning algorithms with the aim of assisting doctors in early disease prediction and diagnosis is very helpful to solve health-related issues. The illness prediction system was created utilizing machine learning methods including the Decision Tree classifier (ML algorithm), Random Forest classifier (ML algorithm), and Naive Bayes classifier (ML algorithm) these are demonstrated in this research project. (Sagarnal, 2020) Tavish Jain, Rohan Sharma, and Ruchika Malhotra proposed a work that illustrates the application of Connectionist Temporal Classification with a CNN-Bi- LSTM model. Convolutional layers for the feature extraction, a Bi-LSTM network for context vector prediction, and a final decoding step that converts each letter in recognizable sequence using the suggested model is composed of LSTM layers into an alphabetic character with the help of CTC loss function. (T. Jain, 2016) Narayana Darapaneni, Richa Agrawal, Arti Kumari, Mohit Gupta, Sachin Padasali, Prabu Purushothaman, and Anwesh Authors have proposed a paper that provides a prediction model basedon easily examined circulatory blood indicators to possibly identify high-risk COVID-19-infected individuals. These findings may be used to develop very effective and much efficient care plans for patients with high risk and routine monitoring for the low-risk patients, improving the patient flow in the hospital. They can also be used to evaluatehospital bed utilization. The current SV-LAR model, which is based on machine learning, recognizes COVID-19- infected patients as high-risk patients who require hospitalization with an 87per f1 score, harmonic mean of 91per accuracy,

and 83per recall. (al, 2021) Dr.P.Hamsagayathri, and Mr .S. Vigneshwaran state in their paper that The development of computer aided diagnostic applications has received a lot of attention recently since poormedical diagnosis can lead to therapies that are seriously misleading. The use of machine learning in computer- assisted diagnostic testing is crucial. As a result, learning from examples is primarily necessary for pattern recognition.Pattern recognition and machine learning (ML) provide the potential of enhancing the accuracy of disease diagnosis and detection in the biomedical field (Vigneshwaran, 2021).

PROPOSED METHODOLOGY

The user interface for the health regulation portal consists of 3 modules follows:

User Module: It serves as a user's main webpage where they may browse hospitals, their costs, and the availability oftreatments and beds at set rates. The user may file a complaint if they become aware of any malpractices. Users can also create queries to track down any price manipulationfraud that may be taking place. Users get access to government-announced policies with their criteria. Users canalso make use of tools like general health predictions, which employ a few algorithms to forecast diseases, and require users to enter correct information about their disease's symptoms in order to get accurate results. They can also use tools like medical prescription recognition, which can detect incorrect medication dosage due to poor handwriting.

Hospital Module: The hospital's module is focused onmaintaining doctors' data & facilities' data. The available treatment for which a doctor is specialized for particular treatment can also be uploaded, so users can make appropriate choices. It consists of login authentication for each hospital. Hospitals can create new data & request approval of data updating to Admin. Admin Module: Admin can take action on Hospital's updating requests based on provided attachments. The administrator has access to every user inquiry regarding pricing manipulation. If a particular Hospital is discovered to have committed a crime or engaged in malpractice, Admin has the full authority to have it removed from the site. Acomponent of the admin module is the acceptance of doctor permission.

In this, we are also implementing machine learning applications in healthcare such as disease prediction and prescription handwriting recognition, for this procedure
is follows as:

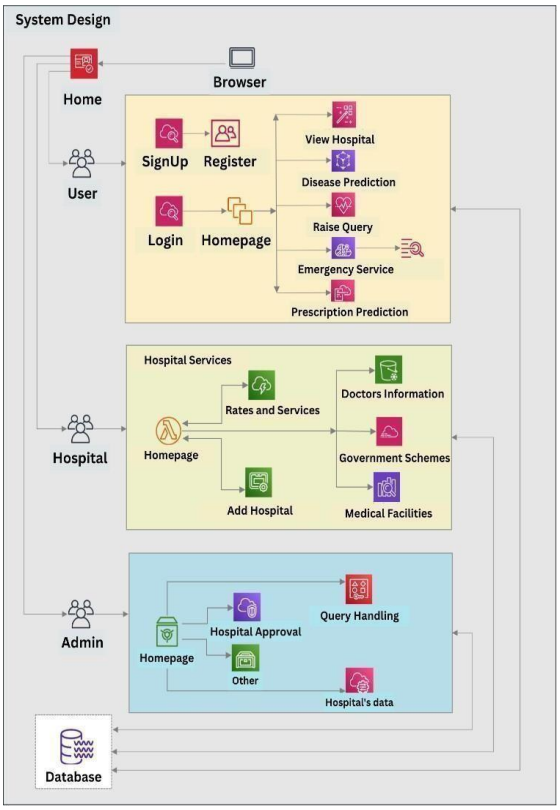

Figure 1: System Architecture

A. Disease prediction:

❖ Input (Symptoms)

We assume that the user is aware of the symptoms he is experiencing. Predictions are developed taking account multiple symptoms, and the input is the user's processing of those symptoms

❖ *Preprocessing Data*

Preprocessing of data involves converting unstructured data into more usable forms. A form that is easily interpreted by an algorithm. In the present work, we used the following preprocessing techniques:

Data Cleaning: During data cleaning process, inconsistenciesin the data are corrected by filling in missing values.

Data Reduction: Analyzing large databases becomes difficult. As a result, we take out the independent factors (symptoms) that might have little or no bearing on our targetvariable (disease). The ratio and outcome of disease- related symptoms are shown in the current study. The trained data sethas 130 columns and 4921 rows.

❖ Model selection:

Three algorithms are used to train the system to predict diseasesas follows -

- ➢ Support Vector Machine
- ➢ Random forest Algorithm
- ➢ Naïve Bayes Algorithm

❖ Equation to calculate accuracy –

$$\frac{Correct\ datapoint\ selected\ by\ algorithm}{Total\ no.of\ datapoints} * 100$$

❖ Algorithms:

These techniques are used to implement the Disease Prediction System. These algorithms' information and operation are provided below.:

Random forest Classifier: This adaptable, user-friendly machine learning method frequently delivers superior results. As stated, overfitting is a significant flaw in the decision tree method. The information seems to have been committed to memory by the tree.

Naive Bayes Algorithm: It is a learning approach whichis supervised. It is used to classify problems on the basis of Bayes theorem. If the tree has retained the knowledge, typically used in text classification tasks with a sizable training set.

Support Vector Machine: SVM is one of the best methods for Supervised Learning, utilized for regression andespecially classification problems. SVM algorithm aims to build the optimal line or set point that can divide spaces intocategories. A hyper-plane is the name for these limits of the correct concept. SVM selects the extreme points/vectors thatcontribute to the creation of the hyper-plane. The term "support vector" refers to such extreme cases, while "support vector machine" refers to the algorithm.

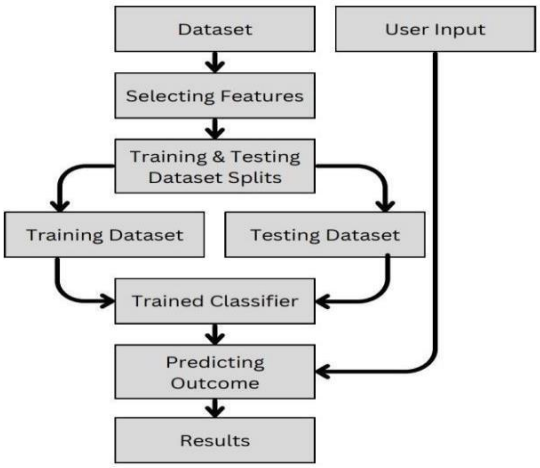

Figure 2: *Processing for Disease Prediction*

❖ Output:

We trained the system by implementing the above-mentioned techniques, with a training dataset. When the user gives input by selecting symptoms into the model, theyare processed further, allowing categorization and prediction to be performed.

B. Medical Prescription Reader

Pre-processing, model-building, processing, and post-processing are all included in a flow. The prescription willfirst be scanned by the model, and only then will preprocessing begin. Resizing, grayscale conversion, noise-reduction, and picture normalization are all parts of the preprocessing stage. Then, feature extraction and classification will be carried out using a convolutional neural network (CNN). The findings will be compared to the acquired dataset in the final step of processing. In order to identify the names of low-accuracy pharmaceuticals by comparing the output to the data set including allmedications, low-accuracy drugs were exposed to the OCRtechnique.

❖ Pre-processing: Prescription photos are first entered into the system through the smartphone camera in PNG format during the pre-processing step. The image is then resized by removing any white areas and turning it to grayscale. Furthermore, implemented a morphological approach. Then the picture is adjusted by comparing the pixel which matches the input image with the neighbors, to equalize the size of all the images. The prescription was then finally divided into three portions using the cropping approach. The header, body, and footer are the three parts. The patient's name, age, weight, and details about the hospital and doctor are all included in the header, which runs from the top to the Rx sign. Next is the body, which starts with the Rx symbol. Since the body contains the quantities of the prescribed drugs, it is the most important component. We place the least importance on the third block in the footer, which contains the doctor's contact information.

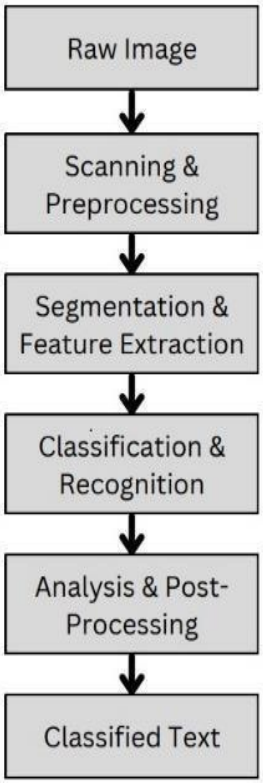

Figure 3: *Processing for Medical Prescription Reader*

❖ Processing: The central component of the data set, which contains the prescription pharmaceuticals, will be classified
when the pre- processing step is finished, and then feature is retrieved using backward and forward convolutions in the convolutional neural network (CNN). CNN accomplishes two tasks— feature extraction and classification—to efficiently identify pictures. As feature extraction methods, Convolutional, Activation Function, and Max pooling layers are used on the input picture. The picture is then produced once the completely linked layer has been applied for categorization. It is made up of several layers that fall under the category of feature extraction and categorization. First, the convolution stage, also known as the feature extraction stage, which calls for the image to be input, a feature detector, and feature map, will be launched. To fill in or complete the feature map, we will next take the center preprocessed picture and apply the filter pixel-by-pixel, block-by-block, by applying a multiplication matrix. Several feature maps are generated in order to obtain our initial convolutional layer. Then, an edge detection filter will be created with the use of the Sobel method. The layer that comes after the convolution layer gives the feature map an activation function to increase the network's nonlinearity. Also, as the max-pooling method makes it simpler to spot and distinguish items throughout the entire image, we use it to gradually lower the input representation size in order to acquire spatial variance. Pooling handle the overfitting issue in addition to reducing processing and the number of parameters required. In order for the data to enter the ANN's input layer and undergo further processing, a sequential long vector is eventually created from the flattened pooled feature map.

❖ Post-Processing: To increase the model's functionality and accuracy, we will continue to gather handwritten
prescriptions for medicines. On the medications that are produced, further classification techniques, such as optical character recognition (OCR), will also be used if the accuracy is 50% or lower. The OCR result will be compared with a data set that contains names of all the pharmaceuticals in order to determine which drug in the dataset is closest to the result. Our main objective in getting the dataset which was gathered from multiple hospitals and doctors with a range of specialties was to gather a large number of individual prescriptions for each medicine that was written in different handwritings.

CONCLUSION

As we previously said, the whole effort is to regulate private healthcare is largely focused on many areas, including search capabilities, technology for prescription detection, customer question resolution in the event of any malpractices, price restrictions for private hospitals, etc. Our primary goal is to reduce patient's discomfort and encourage them to supply critical information with the assurance of a one-stop solution portal. In today's environment, giving information about physicians, therapies, and departments may be a huge game changer for giving steady price ranges & a trustworthy source to go to for treatments. In future, using ML & Data Science & available data of users, we can also predict patterns among patient's diseases, we can avoid data redundancy. As a continuous update, we can make this portal a centralized healthcare portal of a nation where we can book an appointment for doctor, we can integrate facilities of nearby pharmacies in it too.

REFERENCES

1. A. Thamara, M. Elsersy, A. Sherif, H. Hassan, O. Abdelsalam and K. H. Almotairi, Title: "A Novel Classification of Machine Learning Applications in Healthcare," 2021 3rdIEEE Middle East and North Africa Communications Conference (MENACOMM), Agadir, Morocco, 2021, pp.80-85, doi: 10.1109/MENACOMM50742.2021.9678232.
2. E. Hassan, H. Tarek, M. Hazem, S. Bahnacy, L. Shaheen and W. H. Elashmwai, Title: "Medical Prescription Recognition using Machine Learning," 2021 IEEE 11th Annual Computing and CommunicationWorkshop and Conference (CCWC), NV, USA, 2021, pp. 0973-0979, doi: 10.1109/CCWC51732.2021.9376141.
3. Uamakant, B., 2017. A Formation of Cloud Data Sharing With Integrity and User Revocation. International Journal Of Engineering And Computer Science, 6(5), p.12.
4. Butkar, U. (2014). A Fuzzy Filtering Rule Based Median Filter For Artifacts Reduction of Compressed Images.
5. Butkar, M. U. D., & Waghmare, M. J. (2023). Hybrid Serial-Parallel Linkage Based six degrees of freedom Advanced robotic manipulator. Computer Integrated Manufacturing Systems, 29(2), 70-82.
6. Butkar, U. (2016). Review On-Efficient Data Transfer for Mobile devices By Using Ad-Hoc Network. International Journal of Engineering and Computer Science, 5(3).
7. Butkar, M. U. D., & Waghmare, M. J. (2023). Novel Energy Storage Material and

Topologies of Computerized Controller. Computer Integrated Manufacturing Systems, 29(2), 83-95.
8. Butkar, U. (2014). Synthesis of some (1-(2, 5-dichlorophenyl)-1H-pyrazol-4yl (2-hydroxyphenyl) methanone and 2-(1-(2, 5-dichlrophenyl)-1H-pyrazol-4yl) benzo (d) oxazole. International Journal of Informative & Futuristic Research (IJIFR), 1(12).
9. Butkar, U. (2014). An execution of intrusion detection system by using generic algorithm.
10. Butkar, M. U. D., & Waghmare, M. J. (2023). Crime Risk Forecasting using Cyber Security and Artificial Intelligent. Computer Integrated Manufacturing Systems, 29(2), 43-57.
11. K. R. Dalal, Titles: "Analyzing the Implementation of Machine Learning in Healthcare," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 133-137,doi: 10.1109/ICESC48915.2020.9156061.
12. D. Dahiwade, G. Patle and E. Meshram,Title: "Designing Disease Prediction Model Using Machine Learning Approach," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1211-1215, doi:10.1109/ICCMC.2019.8819782.
13. Butkar, U. D., & Gandhewar, N. (2022). Accident detection and alert system (current location) using global positioning system. Journal of Algebraic Statistics, 13(3), 241-245.
14. Butkar, M. U. D., & Waghmare, M. J. (2023). An Intelligent System Design for Emotion Recognition and Rectification Using Machine Learning. Computer Integrated Manufacturing Systems, 29(2), 32-42.
15. Butkar, M. U. D., & Waghmare, M. J. (2023). Advanced robotic manipulator renewable energy and smart applications. Computer Integrated Manufacturing Systems, 29(2), 19-31.
16. Butkar, M. U. D., Mane, D. P. S., Dr Kumar, P. K., Saxena, D. A., & Salunke, D. M. (2023). Modelling and Simulation of symmetric planar manipulator Using Hybrid Integrated Manufacturing. Computer Integrated Manufacturing Systems, 29(1), 464-476.
17. Butkar, U. D., & Gandhewar, D. N. (2022). ALGORITHM DESIGN FOR ACCIDENT DETECTION USING THE INTERNET OF THINGS AND GPS MODULE. Journal of East China University of Science and Technology, 65(3), 821-831.
18. Butkar, U. (2016). Identity Based Cryptography With Outsourced Revocation In Cloud Computing For Feedback Managment System For Educational Institute. Internatinal Journal Of Advance Research And Inovative Ideas in Education, 2(6).