Classification of Spam Emails using Machine Learning and NLP

Prof. D. V. Varaprasad, M.Tech, (Ph.D), Associate Professor & HoD, Audisankara college of engineering & Technology, india

Mrs Uma Maheswari, Assistant Professor, Department of CSE, Audisankara college of engineering & Technology ,india

Gudapareddy Lakshmi, Department of CSE, Audisankara college of engineering & Technology, india

Abstract: Some significant features for automated pill recognition include the colour, size, and form of the pill. But the environment might have an influence that changes the criteria above. Medication mistakes happen a lot and can make things worse for patients. These mistakes happen because of damaged labels, taking the wrong medicine, and other things. This article suggests combining Keras and TensorFlow to train a system that can quickly and easily identify different types of medications. The pill that was found (object detection) links to the pill database, which finds the name of the pill. After the detection phase, the pill is found using the pre-trained dataset. The collection would also include the use cases and specific information about each drug that is needed. The initiative is about gathering datasets for technologies that can automatically find medicines. The experimental findings show that the proposed approach works.

Index terms - — *Pill Identification, Deep Learning, Object Detection, Imprinted Characters, YOLOv5, Convolutional Neural Network (CNN), Pill Classification, Character-Level Language Model, Image Recognition, Keras, TensorFlow.*

1. INTRODUCTION

Over time, problems with medical care have become one of the most common causes of mortality, with an estimated 400,000 deaths per year. The EHRs and Medical Institutions show reports of the medical error epidemic, which shows that the most common type of curable medical error is the drug error. The Institute of Medicine's 2006 study also talks on how to cut down on drug mistakes, which cost a lot of money. From writing a prescription to keeping an eye on how the patient reacts to it, implementation can happen. Maybe the average person doesn't realise how serious medical mistakes may be. For example, if a patient takes the wrong pill (the name and form of the pill), they could end up becoming poisoned by the drug or taking it without needing it. Information extraction from clinical language can help with data mining, finding study participants, managing terminology automatically, de-identifying clinical text, analysing the prescription for an illness and its adverse effects, and other things. Most biological data is usually in an unstructured form because it was transcribed by hand. Without the pill's outer cover, it's hard to tell the chemical makeup and medical name of the medicine, even for someone who is qualified to do so. Most tablets don't have any physical marks that show what they are or what they are made of. It is practically painful for older people, kids, and anybody who isn't used to taking pills to figure out what they are. This makes them take the wrong drug, get the wrong medicine at the wrong time, or obtain the wrong medicine altogether. This negligence can lead to physical side effects or medical poisoning, which can put a patient in the hospital and require a lot of medical treatment before it becomes deadly. We are using deep learning tools like Keras and TensorFlow to help us make a model of each pill. We then feed image descriptions of the pills into the model and train it to figure out which pill is which. This way, we can make an app where the patient shows the model their medicine through their camera. The model will then match the pictures of the pills with the pictures it already has in its data feed, giving the patient the pill's name, composition, and dosage limits. Tensor flow may help you figure out what tablets are by looking at their physical structure and chemical makeup, which tells you how much weight they can hold and how many milligrammes they have in an ounce. Data pre-processing is needed for information extraction since the input data in its narrative form can't be used for summarising or jobs that need decision assistance. In terms of the database, the model needs a lot of pill pictures to learn from. This dataset should have pills of diverse forms, sizes, colours, and markings. You might utilise an already-existing dataset like the Pill Image Recognition Dataset (PIRD) or the Pillbox dataset from the National Library of Medicine.

2. LITERATURE SURVEY

a) AN ACCURATE DEEP LEARNING-BASED PILLDETECTION WITH INTELLIGENT MEDICINAL DRUG IDENTIFICATION SYSTEM

https://www.jmeonline.in/admin/uploadss/Sha_R_P%20(1)%20(3)%20(3).pdf

This study gives a thorough overview of how deep learning methods have been used to identify pills. It looks at how neural networks have changed over time, from early designs to the most advanced models, and talks about how well they can recognise pills based on their form, colour, and markings. The study also talks about problems, such the need for diverse datasets and the need for models to be easy to understand, and it offers ways for future research to go.

b) Pharmacists' and patients' roles in the pharmacist-patient relationship: are pharmacists and patients reading from the same relationship script?

https://www.sciencedirect.com/science/article/abs/pii/S1551741106000489

Background

Pharmacists' jobs have grown to encompass giving information, teaching, and providing pharmaceutical care services. These developments have led to an emphasis on professional interactions between pharmacists and patients that are based on working together, with both pharmacists and patients having roles and duties.

Goal

Using role theory, the study looked at what pharmacists and patients thought about several responsibilities that pharmacists and patients play in the professional connection between pharmacists and patients. Researchers looked at three aspects of the pharmacist-patient relationship: (1) "sharing information," (2) "responsible behaviour," and (3) "interpersonal communication." "Creating a patient-centered relationship" and "active communication related to health care" were two further aspects of the chemist and patient roles that were looked at.

Ways

We got the data by sending questionnaires to 500 randomly chosen patients aged 18 and up and 500 randomly chosen chemists. Using Cronbach's coefficient alpha and bivariate correlation analysis, we figured out how reliable the chemist and patient role dimensions were for internal consistency. We utilised the student's t-test to assess how chemists and patients saw role aspects (alpha level of significance =.05). We employed descriptive statistics to describe the samples of chemists and patients.

Results

The adjusted response rates for the chemist and patient groups were 34.9% (173 out of 496) and 40.8% (196 out of 480), respectively. The reliability coefficients for the chemist and patient role aspects were good. The results showed that pharmacists and patients mostly agree on pharmacists' roles in "information sharing," but patients don't agree as much on pharmacists' roles in "responsible behaviour," "creating a patient-centered relationship," and "interpersonal communication." Patients and chemists don't agree on what patients' duties are in the relationship when it comes to "information sharing," "responsible behaviour," "interpersonal communication," and "active communication related to health care." The results show that chemists are more likely to agree that these are patient duties in the interaction than patients are.

Final Thoughts

If chemists and patients agree on their duties in the interaction, it will work better and have better results. We need to do more study in the future to keep an eye on how chemists and patients feel about their responsibilities in relationships and to come up with new ways to make sure that they are both following the same relationship script.

c) The impact of automation on the safety of drug dispensing in nursing homesImpacto de la automatización en la seguridad de la dispensación de medicamentos a centros sociosanitarios

https://www.sciencedirect.com/science/article/pii/S1130634323004208

Goal

To compare the number and severity of reported medication mistakes in nursing homes that use manual medication dispensing with those that use automated medication dispensing with a well chosen Automated Dispensing System.

Method

A research that looked back at what happened before and after at seven nursing facilities. There were two periods of voluntarily reported dispensing errors: one was in 2013 when weekly pill boxes were dispensed by hand, and the other was in 2015 when an automated drug dispensing and packaging system called Xana 4001U2 Tosho® was used for oral solid medications along with a manual system for other drug forms. We looked at both eras' data on patients' function, cognition, and medications.

Outcomes

The average age of the inhabitants (83.9 vs. 83.6 years; P> .05) and their physical health (Barthel index 41.8 vs. 44.2; P> .05) were similar, but their cognitive health (MMSE 20.3 vs. 21.7; P < .05) was not. There were 408 mistakes found in 2013 (when the system was human), but only 36 faults found in 2015 (when the system was automated). This means that there are 91% fewer mistakes while giving out medicine. In 2013, 43 mistakes got to the patient, whereas only 6 mistakes got to the patient in 2015. Five of these faults needed to be watched, whereas one error did not.

Final Thoughts

The Automated Drug Dispensing and Packaging System makes it more safer to give out and take solid drugs in nursing homes. Voluntarily reporting mistakes made it easier to compare safety between the two times with various dispensing technologies.

d) Implementation of distributed automated medication dispensing units in a new hospital: Nursing and pharmacy experience

https://pubmed.ncbi.nlm.nih.gov/33931903/

Goals and objectives: To look into the structures, procedures, and results of putting in place an Automated Medication Dispensing system and how it affects patient safety.

Background: Over the past 20 years, the process of prescribing, distributing, administering, and managing medicine has become more computerised. Automated medicine dispensing systems are meant to give safe, high-quality, patient-centered treatment, however their use may have unforeseen effects that lead to less than ideal results.

Design: This study takes a qualitative method based on Donabedian's paradigm for structure, process, and result.

Twenty-six registered nurses and pharmacy assistants from clinical areas with automated medicine dispensing cabinets took part in semi-structured interviews. The structures and processes were looked at in detail using theme analysis. Along with interview data, we also looked at text data from internal risk management and critical incident reporting systems to see how well things worked out. The Interactive Sociotechnical Analysis method to health information technology was used to look at the results. This article was written with the help of the COREQ checklist.

Results: Pharmacy assistants were happier with the system when it was first put in place than nurses were. People said that the training nurses got and their role in putting the system in place were not enough. However, over time, nurses' utilisation of and satisfaction with the system got better. Changes made by the system and nurses' innovative problem-solving (workarounds) to deal with these changes had a recursive effect on nurses' productivity and patient safety.

Conclusion: The personalised nature of the "workarounds" used created both dangers and possibilities that need to be further identified, studied, and managed.

Relevance to clinical practice: Nurses make up most of the health care workers. Digitalising health care tasks that used to be done on paper, which affects nursing work, needs the same kinds of methods as any other shift in practice.

e) Analysis of Dimensionality Reduction Techniques on Big Data

https://ieeexplore.ieee.org/document/9036908

A lot of data is being created in many areas, such healthcare, manufacturing, sales, IoT devices, the Web, and organisations, because of digitisation. We employ machine learning methods to find patterns in the different parts of this data. So, medical professionals and individuals in charge may utilise them to generate forecasts that help them make judgements. Not all of the features in the datasets that were created are necessary for training the machine learning algorithms. Some traits may not matter and some may not change the prediction's result. Machine learning algorithms have an easier time when you ignore or remove these less significant or irrelevant features. This study looks at two well-known methods for reducing dimensionality, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA), on four well-known Machine Learning (ML) algorithms: Decision Tree Induction, Support Vector Machine (SVM), Naive Bayes Classifier, and Random Forest Classifier. The study uses a publicly available Cardiotocography (CTG) dataset from the University of California and Irvine Machine Learning Repository. The results of the experiments show that PCA is better than LDA in every way. Using PCA and LDA doesn't change the performance of the classifiers Decision Tree and Random Forest too lot either. We use the Diabetic Retinopathy (DR) and Intrusion Detection System (IDS) datasets to test PCA and LDA even more. Results from experiments show that ML algorithms with PCA work better when the datasets have a lot of dimensions. When the dimensionality of datasets is low, it is clear that ML techniques that don't reduce dimensionality work better.

3. METHODOLOGY

i) Proposed Work:

In this proposed work, a deep learning-based system is developed for the automatic detection and identification of pills using visual features and imprinted characters. The system utilizes YOLOv5, an advanced object detection algorithm, to accurately locate and identify the pill's position in real-time images. Alongside object detection, the system extracts essential visual features such as shape, color, and size. By combining these physical attributes with the detected imprinted characters, the model ensures a more precise identification of pills, even when they are visually similar in appearance.

Furthermore, the proposed model integrates Convolutional Neural Networks (CNNs) and a character-level language model to understand and classify the imprinted text on pills. This hybrid approach improves recognition accuracy, especially in scenarios where pills are deformed, labels are missing, or environmental conditions impact the image quality. The system is trained using a comprehensive dataset consisting of various pill types, including their metadata such as usage, dosage, and side effects, stored in a searchable database. This enhances the efficiency and reliability of the model in real-time pill recognition and medication verification systems.

ii) System Architecture:

The system architecture consists of four primary components: image acquisition, feature extraction, pill detection, and pill identification. First, the input image of a pill is captured and preprocessed to enhance quality and remove noise. The YOLOv5 model is then applied for object detection, accurately locating the pill and segmenting the image. Following this, Convolutional Neural Networks (CNNs) extract visual features such as shape, size, and color, while a character-level language model identifies imprinted characters. These combined features are matched against a centralized pill database to retrieve the corresponding pill name and detailed information. The architecture ensures high accuracy and quick identification, even in complex or noisy environments.

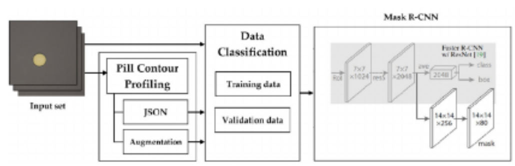

Fig: proposed architecture

iii) Modules:

a. Dataset Upload & Analysis:

In this module, the pill image dataset is uploaded, and preliminary analysis is performed. It includes examining features such as pill shape, color, size, and imprinted characters. The dataset also includes metadata like pill name, use case, and dosage for training and validation.

b. Dataset Processing & Analytical Methods:

This module involves preprocessing images (resizing, noise removal, grayscale conversion) and encoding imprinted character labels into numeric values. The dataset is split into training and testing sets, using 80% for training and 20% for evaluation to enhance model generalization.

c. Run Deep Learning Model:

This module trains the YOLOv5 object detection model and CNN-based classification models on the processed dataset. Imprinted character recognition is also integrated using a character-level language model, and a prediction model is built using these features.

d. Predict Output:

In this module, a new test pill image is uploaded. The trained model detects the pill, extracts features and imprinted characters, and accurately predicts the pill's name and related medical information by matching it against the pill database.

iv) Algorithms:

a. YOLOv5 (You Only Look Once – Version 5):

YOLOv5 is an advanced object detection algorithm that performs detection in a single step, making it fast and efficient. It takes the entire image as input and predicts bounding boxes and class probabilities directly from full images in one evaluation. In this project, YOLOv5 is used to detect pills and their imprinted characters in real-time. It helps in locating the exact position of the pill in the image, regardless of background or lighting variations. Its lightweight architecture and high speed make it suitable for practical deployment on mobile or embedded systems.

b. Convolutional Neural Network (CNN):

CNN is a type of deep learning model specially designed for image-related tasks. It consists of multiple layers such as convolutional layers (for feature extraction), pooling layers (for dimensionality reduction), and fully connected layers (for classification). In this project, CNN is used to analyze pill images and extract important visual features like shape (round, oval), color (white, red, etc.), and surface texture. These features help differentiate between pills that may look very similar but differ in small details. CNNs are effective in learning spatial hierarchies of features from input images.

4. EXPERIMENTAL RESULTS

We tested the suggested method using a large variety of pill pictures that had different sizes, colours, and text printed on them. We divided the dataset into two parts: 80% for training and 20% for testing. The YOLOv5 model was quite good at finding pills and figuring out where the imprinted characters were. The CNN and character-level language model were very good at finding and identifying pill characteristics. The combined technique showed a big boost in prediction accuracy over traditional models, notably for tablets that seem the same but have distinct imprints. The technology was able to find tablets in real time, even when the lighting and backdrop changed.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{ the AP of class } k$$

$$n = \text{ the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$

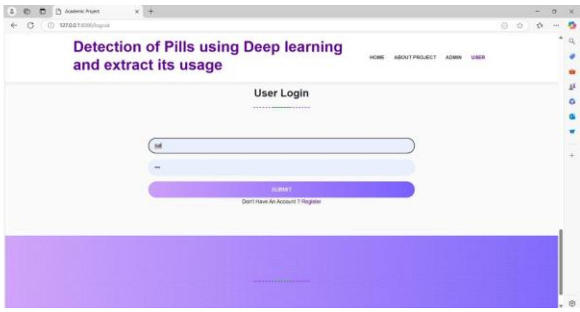

Fig 2: login page

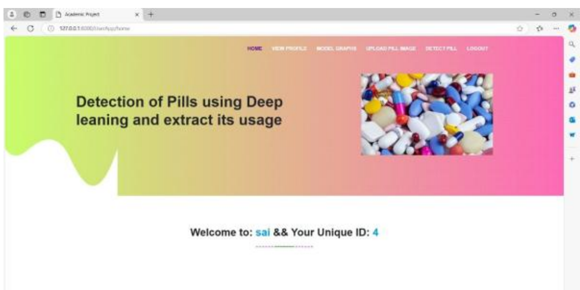

Fig: home page

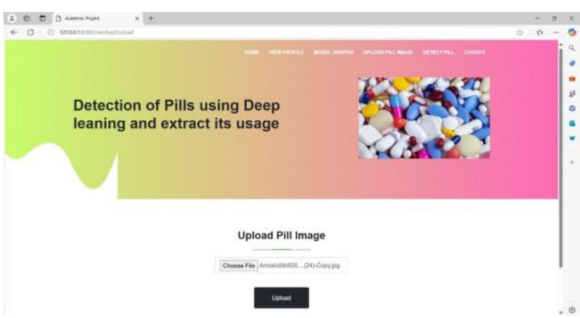

Fig: upload pill image

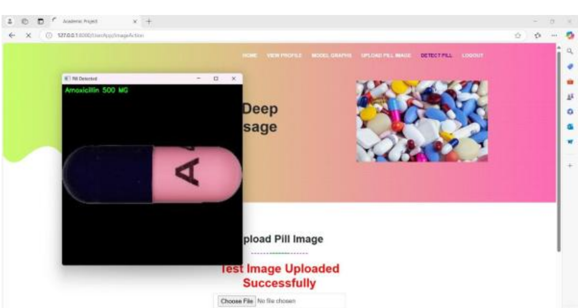

Fig: Pill Name

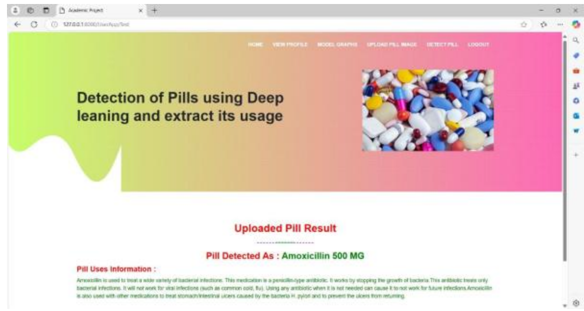

Fig: Pill Description

5. CONCLUSION

The proposed deep learning-based system for pill detection and identification provides an accurate and efficient solution to minimize medication errors. By integrating YOLOv5 for object detection, CNN for visual feature extraction, and a character-level language model for imprint recognition, the system overcomes challenges posed by similar-looking pills and unclear labels. Experimental results show that the model performs well in real-time scenarios, ensuring reliable identification even under complex conditions. This solution has the potential to assist healthcare professionals and patients in verifying medicines safely and quickly.

6. FUTURE SCOPE

In the future, this system can be expanded into a mobile application to provide real-time pill identification using smartphone cameras. This would allow users, especially patients and caregivers, to verify medications easily at home without needing specialized equipment. The application can also support voice search functionality, making it accessible to visually impaired users or those with limited technical skills.

Additionally, the model can be improved by training on a larger, more diverse dataset including regional and international pills. Integration with advanced Optical Character Recognition (OCR) tools can further enhance the detection of imprinted characters, even in low-quality or damaged pills. The system can also be connected to hospital databases and electronic prescriptions, allowing for automatic drug verification, interaction checks, and alerts about incorrect medication intake.

REFERENCES

[1] S. A. Bhatia, "Student Assistant Professor Department of Electronics & communication Engineering, M. Tech, Kurukshetra University (Haryana) HEC Jagadhri (YNR)," IJIRST, Jun. 2016, ISSN.

[2] S. Ramya, J. Suchitra, and R. K. Nadesh, "Detection of Broken Pharmaceutical Drugs using Enhanced Feature Extraction Technique," School of Information Technology and Engineering, VIT University, Vellore, Tamilnadu, India, Apr.-May 2013, pp. 1407. [3] J. O. Gordon, R. S. Hadsall, and J. C. Schommer, "Automated medication-dispensing system in two hospital emergency departments," Am. J. Health Pharm., vol. 62, pp. 1917– 1923, 2005.

[4] E. Y. Fung, B. Leung, D. Hamilton, and J. Hope, "Do Automated Dispensing Machines Improve Patient Safety?" Can. J. Hosp. Pharm., vol. 62, pp. 516–519, 2009.

[5] A. Craswell, K. Bennett, J. Hanson, B. Dalgliesh, and M. Wallis, "Implementation of distributed automated medication dispensing units in a new hospital: Nursing and pharmacy experience," J. Clin. Nurs., vol. 30, pp. 2863–2872, 2021.

[6] G. Pill, "Identification Wizard," Drugs.com, [Online]. Available: https://www.drugs.com/imprints.php, [Accessed: Apr. 13, 2023].

[7] A. Hartl, "Computer-Vision based Pharmaceutical Pill Recognition on Mobile Phones," CESCG, 2010.G. E. Rani, R. Murugeswari and M. Sakthi Mohan, " The innovative secrecy measure for data broadcasting, " 2017 IEEE International Conference on

Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017, pp. 16

[8] G. E. Rani, A. T. V. Reddy, V. K. Vardhan, A. S. S. Harsha and M. Sakthi Mohan, " Machine Learning based Cibil Verification System, " 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 780-782

[9] G. E. Rani, A. T. V. Reddy, V. K. Vardhan, A. S. S. Harsha and M. Sakthi Mohan, " Machine Learning based Cibil Verification System, " 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 780-782

[10] Rani, G.E., Murugeswari, R., Siengchin, S., Rajini, N., & amp; Kumar, M. A. (2022). Quantitative assessment of particle dispersion in polymeric composites and its effect on mechanical properties. Journal of Materials Research and Technology, 19, 1836–1845. [11] G. E. Rani, R. Murugeswari and N. Rajini, " Edge Detection in Scanning Electron Microscope (SEM) Images using Various Algorithms, " 2020 4th International Conference on Intelligent Computing and Control Systems 2020, pp. 401-405

[12] G. Elizabeth Rani., H. Mohan, B. Kusuma, P. S. Kumar, A.M. Jenny and N. Akshat," Automatic Evaluations of Human Blood Using Deep Learning Concepts, " 2021 6thInternational Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 393-396

[13] R. K. Devi and G. Elizabeth Rani, " A Comparative Study on Handwritten Digit Recognizer using Machine Learning Technique, " 2019 IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development (INCCES), 2019, pp. 1-5

[14] M. Sakthi Mohan, P. G. K. Reddy, T. Narendra, B. Venkatesh and R. G. Elizabeth,

" Leaf Health Monitoring and Disease Detection Using Image Processing, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in

Engineering (ICACITE), 2022, pp. 773-777

[15] S. M, G. K. Reddy, T. J. Reddy, B. Manikanta and E. R. G, " Contactless Covid19 Monitoring System Using IOT, " 2022 2nd International Conference on

AdvanceComputing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 670674

[16] E. R. G, S. M, L. C. G. P, A. S. S, S. P and N. Kumar Reddy, " Scam Recognition in Visa/Credit Card Using Genetic Algorithm, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 812816.

[17] S. M, E. R. G, M. Devendra Reddy, D. V. V. S. S. S. Babu, M.V. Vardhan and K. Karthigadevi, " Forecast of Heart Sickness using Machine Learning, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 1115-1119 [18] G. E. Rani, S. M, M. P. Suresh, M. Abhiram, K. J. Surya and B. Y. A. N. Kumar, " Face Recognition Using Principal Component Analysis, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 932-936

[19] M. Sakthi Mohan, G. Elizabeth Rani, S. K. Jeevan Swetha, G. Dharani, K. M. Nikhila and R. Kannigadevi, " An automated face mask detection using machine learning techniques, " 2022 2nd International Conference on Advance Computing and

Innovative Technologies in Engineering (ICACITE), 2022, pp. 899-904

[20] G. Elizabeth Rani, G. Narasimha Murthy, M. Abhiram, H. Mohan, T. Singh Naik and M. Sakthi Mohan, " An Automated Airlines Reservation Prediction System Using

Blockchain Technology, " 2021 Sixth International Conference on Image Information Processing (ICIIP), 2021, pp. 224-228

[21] G. E. Rani, E. Venkatesh, K. Balaji, B. Yugandher, A. Kumar and M. Sakthi Mohan, " An automated prediction of crop and fertilizer disease using Convolutional Neural Networks (CNN), " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 1990-1993.

[22] E. R. G, S. M, A. R. G, S. D, T. Keerthi and R. S. R, " MNIST Handwritten Digit Recognition using Machine Learning, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 768-772.

[23] E. R. G, S. M, R. R. R, S. G. M, S. S. R and K. K, " An Automated Cost Prediction in Uber/Call Taxi Using Machine Learning Algorithm, " 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 764-767

[24] A. Konda and L.C. Xin, "Evaluation of Pilling by Computer Image Analysis," Journal of the textile Machinery Society of Japan, vol. 36, pp. 96-107, 1990.