Medical Application for Leukemia Patients using Deep Learning and Steganography

Abstract

Acute Lymphoblastic Leukemia (ALL) affects more than 500,000 people yearly all across the world. Highly accurate detection of leukemic cells in blood stream without any chemical process will help the healthcare industry not just with cost, but will also save time as the outcome can be processed in a fraction of time. The project aims to build an open-sourced web application specifically meant for inferring whether the microscopic cell imaging indicates the presence of leukemia blasts. The system would request for the microscopic image, process it and pass it to a pre-modelled Deep Learning algorithm to get prediction probability of Leukemia blast. The information would then be hidden in the said image and returned to the patient (so that the patient does not get to know). After the patients passes the image to the doctor, the doctor can use this image to manually infer, and also pass the image to this application and get prediction probabilities to back their claim and have a better understanding.

1. Introduction

Acute lymphoblastic leukemia (ALL) is a type of cancer of the blood and bone marrow — the spongy tissue inside bones where blood cells are made. The word "acute" in acute lymphoblastic leukemia comes from the fact that the disease progresses rapidly and creates immature blood cells, rather than mature ones. The word "lymphoblastic" in acute lymphoblastic leukemia refers to the white blood cells called lymphocytes, which ALL affects. Acute lymphoblastic leukemia is also known as acute lymphocytic leukemia. Newly developed deep learning models have made significant progress in the classification of medical images in various diseases such as diabetic retinopathy, breast cancer, pneumonia, etc. Active research and building of fine-tuned models enable us to unveil unknown details from the pictures and accurate disease prediction. The aim of this project is to create an end-to-end web application that is purely based on concepts of Image Processing such as Image Classification and Information Hiding in Images for medical purposes so as to aid medical practitioners and doctors (oncologists) towards the classification of Acute Lymphoblastic Leukemia in blood cell micro images of the patient. The web application has two main group of users: Patients and Doctors.

1.1 Patients

The patients would upload their medical image to this system and after some processing, they will get a new image in return which will be free of staining noise and illumination errors (if any). Their uploaded image will be passed by a Deep Learning model that will be made using appropriate architectures and the prediction probability generated from this model will be masked into this image using steganography concepts. The final image can be downloaded by the patient and can be then presented to their doctors.

1.2 Doctors

The doctor, after receiving a clearer image from the patient, will upload this image to this system, and the hidden information will be displayed to the doctor (the prediction probabilities). The doctor can come to a conclusion by manually trying to understand the image, and also taking the Deep Learning model's prediction into account.

2. Literature Study

1. Steganography algorithm to hide secret message inside an image

This paper sheds light on a new steganography algorithm that will hide data inside an image more effectively. Binary codes and the pixels in a picture are used in the suggested approach. To maximise the amount of data that may be stored inside an image, the zipped file is employed before the image is converted to binary codes. The proposed algorithm is used to create a system known as the Steganography Imaging System (SIS). The algorithm is then tested on the system to determine its viability. The photos contain data of various sizes, and for each of the examined images, the PSNR (Peak signal-to-noise ratio) is also recorded. The steganographic image has a greater PSNR value than the other images.

2. Classification of acute lymphoblastic leukaemia using deep learning

When left untreated, acute leukaemia, a serious illness that affects both children and adults, can be fatal. Acute lymphoblastic leukaemia (ALL) strikes youngsters quickly and kills them within a few weeks after diagnosis. The haematologists examine bone marrow and blood to determine ALL. Techniques for manual blood testing that have been in use for a long period are frequently slow and produce less reliable diagnoses. With the use of a computer-aided method that uses deep learning and image processing, our work enhances the diagnosis of ALL. The classification of ALL into its subtypes and reactive bone marrow (normal) in stained bone marrow pictures was proposed in this study. Convolutional neural network deep learning

techniques are employed to train the model on bone marrow pictures, which yields accurate classification results. Thus, experimental results were produced and contrasted with those of Naïve Bayesian, KNN, and SVM classifiers. According to experimental findings, the suggested approach obtained 97.78% accuracy. The acquired results show that the suggested strategy may be employed as a diagnostic tool for acute lymphoblastic leukaemia and its subtypes, which will undoubtedly be helpful to pathologists.

3. Convolution Neural Network (CNN) for image detection and recognition

Deep Learning algorithms are created to closely resemble how the human cerebral cortex works. These algorithms are models of deep neural networks, which are neural networks with numerous hidden layers. Convolutional neural networks are deep learning algorithms that can process enormous datasets with millions of parameters as 2D images as input and outputs. In this article, CNN models are developed to assess how well they perform on datasets for image recognition and detection. On the MNIST and CIFAR-10 datasets, the method is put into practise, and its performance is assessed. Models using MNIST are 99.6% accurate, and CIFAR-10 uses real-time data augmentation and CPU unit dropout.

4. Hyperparameter optimization for machine learning models based on Bayesian optimization

Hyperparameters are crucial for machine learning algorithms. They directly influence the actions of training algorithms and have a major impact on the effectiveness of machine learning models. As a result, an effective hyperparameter optimization algorithm can be created to optimise any particular machine learning method. In this study, the authors take into account Gaussian processes to construct the relationship between the performance of machine- learning models and their hyperparameters. Bayesian optimization reduces the tuning problem for machine learning models to an optimization problem. Bayesian optimization is founded on the Bayesian theorem and involves taking information from previous samples. On common test datasets, several experiments have shown that the suggested method can find the optimal hyperparameters for popular models like neural networks and random forest algorithms.

5. Medical image processing, analysis and visualization in clinical research

Imaging has become an essential component in many fields of medicine and science. Analysis of these diverse image types requires sophisticated computerized quantification and visualization tools. These days, a cheap desktop computer outfitted with the necessary graphics

hardware and software can do a lot of the visualisation and analysis. This paper proposes a general-purpose, extensible, platform-independent image processing and visualisation tool made primarily for an online medical research community. Clinical and quantitative analysis of medical pictures can be done online with the help of the MIPAV (Medical Image Processing, Analysis and Visualization) program. Researchers and clinicians at distant sites can readily share research data and analyze using MIPAV's standard user interface and analysis tools, improving their capacity to explore, diagnose, monitor, and treat medical problems.

3. Problem Statement and Objectives

2.1 Problem Statement

Acute Lymphoblastic Leukemia has a very high mortality rate. Building fast and efficient methods of classification of medical diseases is highly essential for doctors, owing to the fact that patients are usually highly vulnerable to facing death situations. Not having efficient open sourced deep learning technologies for doctors and medical specialists to rely on is main problem addressed by this project.

2.2 Objective

To create a web-based application that aims at serving patients suffering from Acute Lymphoblastic Leukemia to aid them with their medical diagnosis using the micro cell imaging of the patients' blood. The application aids doctors to understand medical imaging by providing prediction probabilities generated by a pre-modelled learning algorithm based on a dataset.

4. Design

4.1 Tools Used

• Python
• Django
• HTML
• CSS
• Tensorflow
• Keras
• Pandas
• Numpy
• Open CV

4.2 Architecture Diagram

The diagram below explains the three main areas of focus in the project: The User Interface, the Deep Learning modelling, and the Steganography concept.
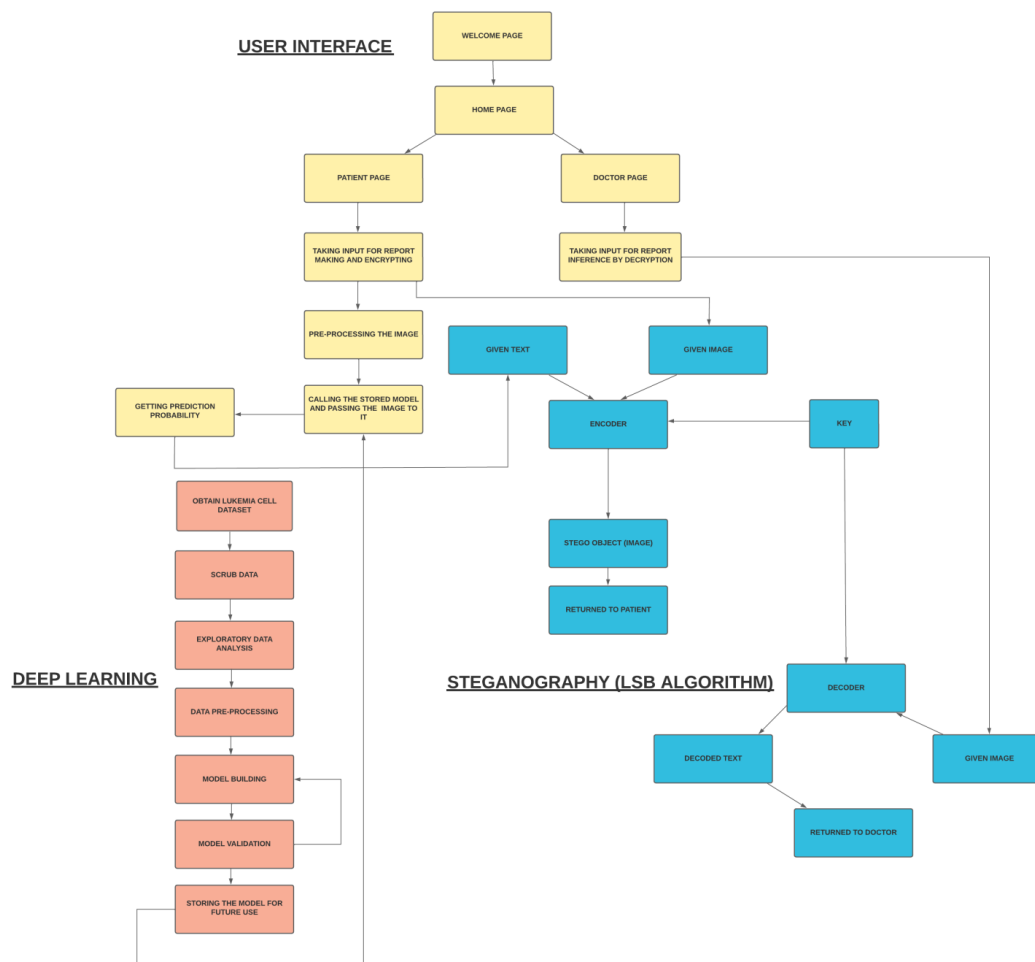

Figure 1.   Architecture Diagram

5. Implementation

5.1 Dataset Details

There are 15,135 images from 118 patients with two labelled classes:

• Normal cell (HEM)

• Leukemia blast (ALL)

Here is how some of the HEM and ALL cells look like:
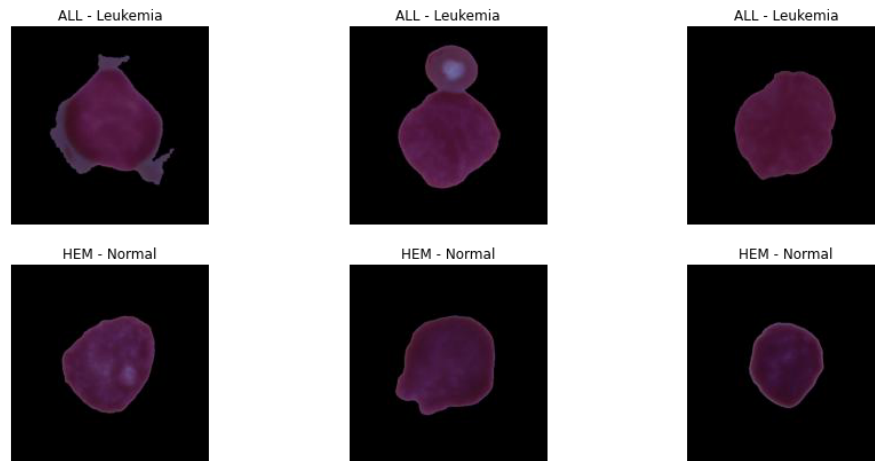
*Figure 2.   Micro Cell Imaging of Blood Cells*

5.2 Exploratory Data Analysis

To understand whether the dataset is imbalanced or not, we use a bar chart to visualize the count of HEM and ALL images, given below:

Figure 3.   Bar Chart presenting imbalanced Data

The images shown in figure 1. has a lot of black background which may not be required for model building. Appl... The ... ldful, and can build better models.
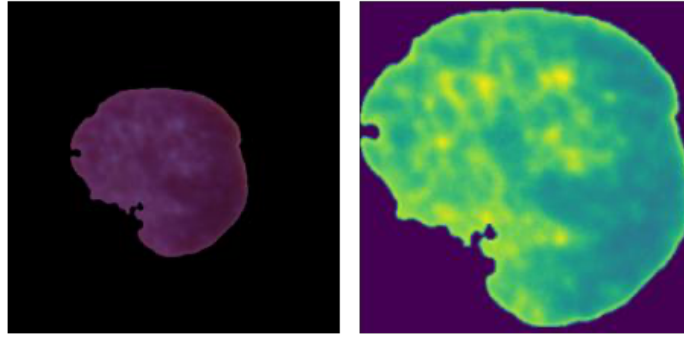
Figure 4.   Cell Imaging before and after Otsu Thresholding

To this threshold image, we fit the deep learning model after processing the dimensions of each image. The model built is a sequential model including 2D-convolution layers along with 2D Max Pooling layers. Dropout layers have been used in places to drop random connections between two layers so that the model does not overfits. Once flattened to a single dimension, Dense layers are used to get a single output value. There are more than 5 million parameters used to train the model. Here, 1 stand for ALL and 0 stands for HEM. The final output is a value between this range. This model is fitted on the training data and validated on the validation data for 50 epochs with a batch size of 32 using adam optimizer and calculating loss using binary cross-entropy. The evaluation metrics are accuracy and recall.

```
Model: "sequential"
_____
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 200, 200, 32)      320

 conv2d_1 (Conv2D)           (None, 200, 200, 64)      18496

 max_pooling2d (MaxPooling2D  (None, 100, 100, 64)     0
 )

 dropout (Dropout)           (None, 100, 100, 64)      0

 conv2d_2 (Conv2D)           (None, 100, 100, 64)      36928

 max_pooling2d_1 (MaxPooling  (None, 50, 50, 64)       0
 2D)

 dropout_1 (Dropout)         (None, 50, 50, 64)        0

 conv2d_3 (Conv2D)           (None, 50, 50, 128)       73856

 max_pooling2d_2 (MaxPooling  (None, 25, 25, 128)      0
 2D)

 dropout_2 (Dropout)         (None, 25, 25, 128)       0

 flatten (Flatten)           (None, 80000)             0

 dense (Dense)               (None, 64)                5120064

 dropout_3 (Dropout)         (None, 64)                0

 dense_1 (Dense)             (None, 1)                 65


=================================================================
Total params: 5,249,729
Trainable params: 5,249,729
Non-trainable params: 0
_____
```

Figure 5. CNN Architecture used

At the end of the training, the model achieves a training accuracy of 99.07% and recall of 99.51% with 0.0287 loss. The validation data achieves an accuracy of 69.01% and recall of 95.63% with 3.7698 loss. The model is saved as an H5 file for it to be loaded later and used for prediction.

A class is created for building the Steganography algorithm. In the Lowest Significant Bit algorithm, the information is hidden in the last bits of the pixels in the cover image. The LSB is the least significant bit in the byte value of a pixel in the image. In a 24-bit image, the three basic colors present that contribute to 24 bytes (8 + 8 + 8) are Red, Green and Blue. Each represents 1 byte. An 800 x 600 image can store 1440000 bits of encrypted information consequently. A change in the LSB of a pixel is evidenced by minor changes in the intensity of the colors. These changes are usually too little to be detected by the naked eye, and thus the steganogramme is generated.
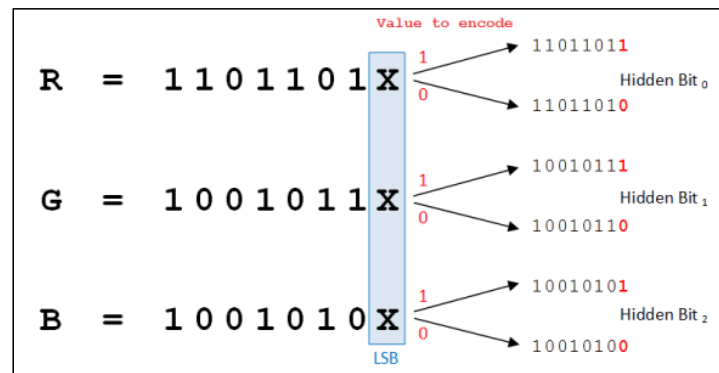
The website is made using Django framework. The home page will provide access to both patients and doctors.
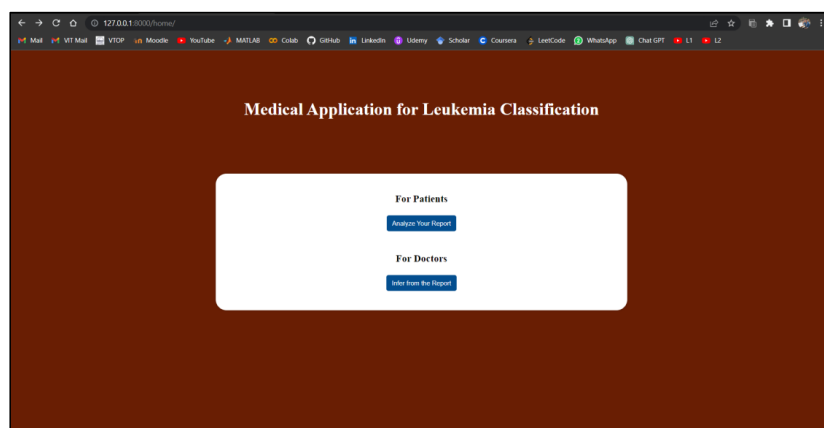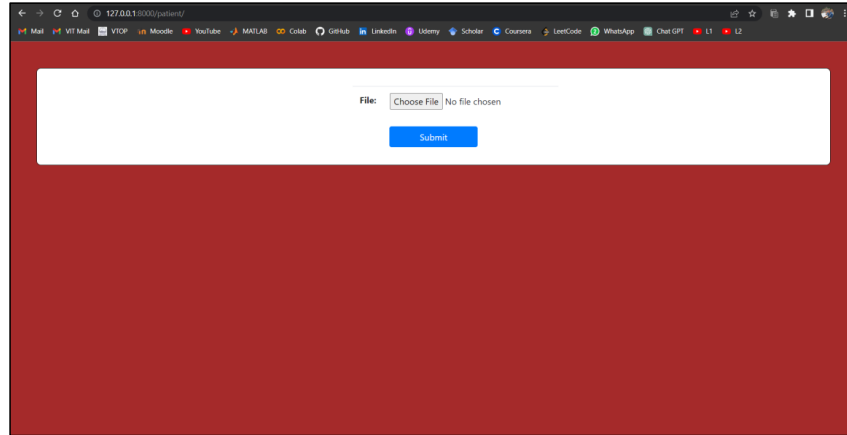

Figure 7. Home Page

On clicking the portal to patients, the user will be redirected to a page where they can upload their medical image. On upload, the trained model will be loaded, the image will be processed as explained above (Otsu Thresholding), its dimensions will be expanded, and then the model will predict the outcome. This outcome will be encrypted into the image and the encrypted image will be returned to the patient for download.

The patient can provide this encrypted image to their doctor. A doctor trying to access the doctor portal from the home page will be redirected to a page where they can upload the image and decrypt it t
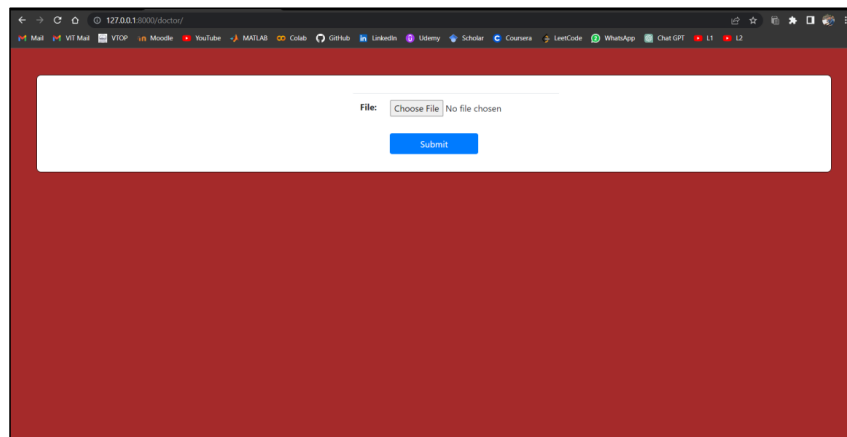
Figure 9. Doctor Page

On decryption, the prediction probability of the image being classified as Acute Lymphoblastic Leukemia is being reflected on the page.

6. Conclusion

The aim and objective of the project has been successfully achieved. An-end-to-end functioning website has been created that takes leukemia cell image from the patient and uses the trained CNN model to classify it as HEM or ALL with a particular prediction probability. This probability values gets hidden in the image using LSB Algorithm and the image is returned

to the patient. When the patient sends this image to the doctor, the doctor can use this site to find out the hidden data by decrypting the image and getting the prediction probability. The doctor can then use their intelligence and the model's prediction to act accordingly. The future work can include training the model for more number of epochs, adding identity access and management to the website, and deploying the website on the internet for everyone to use.

References

[1] Ibrahim, R., & Kuan, T. S. (2011). Steganography algorithm to hide secret message inside an image. arXiv preprint arXiv:1112.2809.

[2] Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. Microscopy Research and Technique, 81(11), 1310-1317.

[3] Chauhan, Rahul & Ghanshala, Kamal & Joshi, R.. (2018). Convolutional Neural Network (CNN) for Image Detection and Recognition. 278-282. 10.1109/ICSCCC.2018.8703316.

[4] Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, Si-Hao Deng, Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization, Journal of Electronic Science and Technology, Volume 17, Issue 1,2019, Pages 26-40, ISSN 1674-862X.

[5] M. J. McAuliffe, F. M. Lalonde, D. McGarry, W. Gandler, K. Csaky and B. L. Trus, "Medical Image Processing, Analysis and Visualization in clinical research," Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001, Bethesda, MD, USA, 2001, pp. 381-386, doi: 10.1109/CBMS.2001.941749.