The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345

Progressive Analysis and Predictions of Leukemia (Cancer) Patients ɑna Machine Learning Model- The APPOLLO Hospitals Accredited

[1]Dr.Kamel Alikhan Siddiqui, [2]Mrs.Kaneez Fatima, [3]Ali Hasan Khan, [4]Mr. Sibghatullah
[1,2,3,4]Lords Institute of Engineering and Technology(Autonomous)

Abstract
This abstract describes a machine learning model for predicting the prognosis of cancer patients with leukemia. The model is based on analyzing patient data such as age, gender, diagnosis, lab results, treatment history, and prior hospitalizations. The model employs supervised learning techniques such as random forests and gradient boosting to generate predictions of survival probabilities. Additionally, the model uses feature engineering to identify important features and reduce noise in the data. The model is evaluated using the area under the receiver operating characteristic curve (AUC) and other metrics. The results indicate that the model has good predictive accuracy and can be used to identify high-risk patients and guide clinical decisions. The goal of this project is to develop a machine learning model for the diagnosis and prognosis of leukemia in Indian patients. The model will use a progressive analysis of patient data to identify the characteristics associated with leukemia, including genetic markers, environmental exposures, lifestyle factors, and demographic information. Using this information, the model will be used to accurately predict the risk of developing leukemia in Indian patients. Additionally, the model will be used to identify the most effective treatments for those diagnosed with leukemia and to monitor the disease progression. Finally, the model will be evaluated for its accuracy and effectiveness in a clinical setting.

Keywords: Leukemia, Cancer in Sub Continent, Analysis on ML Model, Analysis-AUC, Accuracy and Python Programing.

1) **Introduction**:

Cancer is a serious medical condition that affects millions of people around the world. With advances in medical technology and software tools, there are now many innovative ways to provide cancer patients with the best possible care. In the modern age, software tools can help to enhance the quality of care for cancer patients by providing them with access to the latest treatments and research. For example, software tools can be used to track patient data, monitor treatments, and provide personalized care plans tailored to the individual patient's needs. Additionally, software tools can support patient- centered care by enabling patient education and communication with their healthcare providers. With these tools, cancer patients can easily access the latest information about their diagnosis, treatment options, and support resources. Unfortunately, despite these advances, many cancer patients in India and the Subcontinent are still denied access to the best medical resources due to financial constraints. A lack of access to proper medical care, technology, and resources can be detrimental to the health and wellbeing of cancer patients. Furthermore, the cost of treatments and medications can be a major burdenfor many families, leading to inadequate care and, in some cases, a lack of access to life-saving treatments. To address this issue, organizations and individuals have taken steps to support cancer patients in India and the Subcontinent. These efforts include fundraising campaigns, providing free or low-cost treatments, and providing educational

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345
resources about cancer and its treatments. Additionally, various organizations have developed software tools to help health care providers in India and the Subcontinent better manage their patients' care. By providing access to the best medical resources, these programs can significantly improve the quality of life for cancer patients in India and the Subcontinent.

Cancer treatments in India are often more expensive than in other countries because the cost of medical care is relatively high in India. Additionally, many of the treatments available in India are more advanced than treatments available in other countries, which can drive up the cost. Additionally, many of the cancer treatments available in India are not covered by insurance, so patients have to pay out of pocket for the entire cost of their treatment. Machine Learning tools can help diagnose medical conditions in India by providing a more accurate diagnosis and helping to reduce the amount of time it takes to diagnose a patient. Machine Learning tools can help identify patterns in the patient data that can lead to a more accurate diagnosis. For example, ML algorithms can be used to compare patient information (such as age, gender, medical history, etc.) with data from similar patients that have been diagnosed with a particular condition and use this information to provide a more accurate diagnosis. Machine Learning can also be used to identify potential treatments that could benefit the patient and reduce the amount of time it takes to come up with a treatment plan. Additionally, ML algorithms can be used to detect and monitor diseases in the population by analyzing the health data of a large group of individuals and looking for patterns that could indicate the presence of a particular disease.

2) Background Studies:

Leukemia is a type of cancer that affects the blood cells and bone marrow. It is the most common type of cancer in children and young adults in India. It is caused by an abnormality in the production of blood cells and can lead to anemia, infections, and bleeding. It can also spread to other parts of the body, such as the lungs, brain, and lymph nodes. The exact cause of leukemia is not known, but it is believed to be associated with genetic and environmental factors. Possible causes include exposure to radiation, certain chemicals, and certain viruses. There are also some factors that can increase the risk of developing leukemia, such as family history, certain genetic conditions, and certain medications. Early diagnosis and prompt treatment are important for successful treatment of leukemia. Treatment options include chemotherapy, radiation therapy, stem cell transplant, and targeted therapy. In India, access to healthcare and cost of treatment are major barriers to the successful management of leukemia. Leukemia awareness and early diagnosis is essential for effective management and prevention of the disease. Organizations such as the Leukemia and Lymphoma Society of India and Leukemia Care India are working to raise awareness about the disease and provide resources for patients.

Major Issues for Leukemia Patients:

1. Access to Treatment: Access to treatment is an ongoing issue for leukemia patients in India. Many patients suffering from leukemia face difficulty in accessing treatment due to the high cost of chemotherapy and other treatments.
2. Lack of Awareness: Leukemia is an often misunderstood disease in India. There is a lack of awareness among the people about the signs and symptoms of the disease. This leads to delayed diagnosis and treatment.
3. Diagnosis: Diagnosing leukemia can be difficult as the symptoms are often similar to other common illnesses. This leads to delayed diagnosis which can be dangerous for the

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345
patient.

4. Limited Resources: India has limited resources for the treatment of leukemia. There is a lack of specialized centers for the treatment of this disease. Most of the treatment facilities are located in big cities and are often inaccessible for patients from rural areas.
5. Shortage of Blood: Blood transfusions are often required for leukemia patients, but there is a shortage of blood in India. This often results in delay in treatment and can be fatal for the patient. Indian health care facilities are much more affordable compared to other countries.

The exact time it will take for a patient to detect and receive information about Leukemia in an Indian hospital depends on several factors, including the type of test used and the availability of resources. Generally, it can take anywhere from a few days to weeks to get back test results. The most common test used to diagnose Leukemia is a blood test, which looks for abnormal levels of white blood cells. Depending on the hospital, other tests such as a bone marrow biopsy and/or imaging tests of the body may also be used. Once the diagnosis is confirmed, further tests may be required to determine the type of Leukemia and the best treatment options.

White blood cells (WBCs) play an important role in the growth and spread of cancer in leukemia patients. WBCs produce abnormal proteins that help cancer cells to grow and spread. These abnormal WBCs also can prevent the body's natural defenses from fighting off cancer cells. Abnormal WBCs can also promote the growth of new blood vessels, which can help cancer cells to spread and grow. In addition, abnormal WBCs can produce substances that can change the environment of the body and make it more hospitable for cancer growth. Finally, abnormal WBCs can produce substances that can stimulate the growth of new cancer cells. Thus, WBCs are important for the growth and spread of cancer in leukemia patients.

Machine learning techniques can be used to make medical diagnosis and treatments more accurate andefficient.

1. Automated Diagnosis: Machine learning algorithms can be used to automatically diagnose diseases based on patient symptoms, medical scans and other data. This could help reduce the time it takes for doctors to make accurate diagnoses and provide more accurate results.
2. Drug Development: Machine learning algorithms can be used to identify new drugs and treatments that can be used to treat specific diseases. This could help speed up the drug development process andlead to more effective treatments.
3. Personalized Medicine: Machine learning algorithms can be used to tailor treatments to individual patients based on their specific needs and conditions. This could help doctors provide more personalized care and yield better results.
4. Disease Prediction: Machine learning algorithms can be used to predict the likelihood of certain diseases in patients based on their data and medical history. This could help doctors identify patients who are at risk of developing certain diseases and take preventive measures to reduce their risk.
5. Clinical Decision Support: Machine learning algorithms can be used to provide doctors with assistance in making decisions about treatments and medications. This could help reduce medical errorsand provide better patient outcomes.

Automated Diagnosis Machine Learning model:

is a type of artificial intelligence that uses algorithms to interpret data from patient medical

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345
records and other diagnostic tests to automatically diagnose a patient's health condition. It then uses the data to provide an accurate diagnosis, often with a higher accuracy than that of traditional diagnostic systems. The model utilizes algorithms to identify patterns in data and make diagnoses without the need for manual input from a clinician. This model can be trained on a dataset of patient data, such as medical images, vital signs, and lab results. After training, the model can then be used to automatically diagnose new cases. The model is designed to learn from the data, becoming more accurate over time, making it a powerful tool for automating diagnosis.

The model typically incorporates supervised and unsupervised learning techniques, such as Bayesian networks, decision trees, neural networks, and k-nearest neighbor algorithms, to interpret the data and process it in a way that can lead to an accurate diagnosis. Additionally, the model can be further optimized with data augmentation and hyper parameter optimization methods to improve its accuracy and performance.

Drug Development ML Model: Machine learning algorithms Drug Development Machine Learning algorithms are computer programs that use statistical techniques to identify patterns in data and make predictions based on those patterns. These algorithms can be used to identify potential compounds for drug development, predict how likely a drug is to be successful, and identify potential side effects of a drug. They can also be used to analyze historical data to generate trends and insights that can help inform decisions in drug development. Machine learning algorithms use a variety of techniques, such as supervised and unsupervised learning, to analyze large datasets and produce accurate predictions.

Personalized Medicine Model: The most suitable machine learning model for personalized medicine is a supervised learning model. This type of model is capable of predicting medical outcomes based on a variety of factors, including patient history, genetics, lifestyle, and other health-related data. Supervised learning models can be used to develop predictive models that can help identify high-risk patients and suggest specific treatments or interventions. This type of model is also useful for tracking patient outcomes over time and identifying areas where further research is necessary.

A Disease Prediction: Machine Learning Model is a type of artificial intelligence (AI) system that uses data-driven algorithms to identify patterns and correlations in a dataset in order to make predictions about the likelihood of a certain outcome, such as a disease. The model is trained with historical data, such as patient medical records, symptom information, and lifestyle habits. It then uses this data to identify patterns and build a statistical model. The model can then be used to predict the probability of a patient developing a certain disease or illness. The model can also be used to suggest tests or treatments that may be beneficial for a particular patient.

A Clinical Decision Support Disease Prediction Machine Learning Model: is a computer-based system that is designed to help healthcare providers make better decisions when it comes to diagnosing and treating medical conditions. It uses predictive analytics and machine learning algorithms to analyse large amounts of data to identify patterns and trends that could lead to the accurate diagnosis and treatment of diseases. The model can be trained to recognise patterns in medical records, lab test results,imaging studies and other data sources to predict the most likely diagnosis or treatment path. It can also provide alerts and reminders to help healthcare providers stay up-to-date on the latest evidence-based clinical guidelines. This model can help healthcare providers improve the accuracy and speed of diagnosis and

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345
treatment of diseases, leading to better patient.

3) Indian Live Habits:

Indian Live Habits for cancerous leukemia Cancerous leukemia is a rare and serious form of cancer that affects the body's white blood cells. It is a life-threatening disease and requires intensive treatment,including chemotherapy and radiation therapy. In India, the most common form of cancer is leukemia, and it affects children, adolescents, and adults alike. The Indian government has taken steps to bring awareness and provide support to those affected by leukemia. An estimated 1,00,000 new cases of leukemia are diagnosed in India every year. Supportive services like counseling and palliative care are available in select hospitals, and some organizations have been working to support those affected by leukemia. In terms of lifestyle and habits, it is important for those with leukemia to take measures to reduce their risk of infection. This includes avoiding contact with people who have colds or flu, washing hands frequently, and eating a healthy, balanced diet. It is also important to get plenty of rest and exercise, and to limit stress. In terms of treatment, many Indian doctors recommend a combination of both Western and traditional Indian medicines. This may include chemotherapy, radiation therapy, and Ayurvedic treatments. Ayurvedic treatments involve herbs and herbal concoctions, and many believe that these treatments can help reduce the side effects of chemotherapy and radiation therapy. Finally, itis important for those with leukemia to take measures to protect their mental health. This includes connecting with supportive family and friends, seeking out support groups, and staying positive.

4) Proposed Design:

There are several machine learning models that can be used for progressive analysis of leukemia patients. The choice of model depends on the specific problem and data available. This paper involvedin analysis of the following Machine Learning models.

1. Decision Trees: Decision trees are useful for classification problems and can be used to predict whether a leukemia patient will progress to a more severe stage based on specific features.
2. Random Forest: Random forest is a type of ensemble learning method that combines several decisiontrees to make more accurate predictions. It can be used to classify leukemia patients into different riskgroups based on various clinical and genomic features.
3. Support Vector Machines (SVM): SVM is a classification algorithm that separates data into differentclasses using a hyperplane. It can be used to classify leukemia patients based on different clinical and genomic features.
4. Artificial Neural Networks (ANN): ANN is a type of deep learning algorithm that can learn from large amounts of data to make accurate predictions. It can be used to predict the progression of leukemiabased on various clinical and genomic features.

It's important to note that the performance of these models will depend on the quality and quantity of data available, as well as the specific features used to train the model. It is recommended to consult with a healthcare professional or a data scientist with experience in medical applications of machine learning before making any clinical decisions based on the output of these models.

5) Machine Learning intuitions:

1) Using Decision Trees:

python codes for Automated Diagnosis Machine Learning Model using Decision trees

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345
#Importing the necessary libraries

import pandas as pd
from sklearn.tree import DecisionTreeClassifier from sklearn.model_selection import

train_test_splitfrom sklearn.metrics import accuracy_score #Loading the dataset

data = pd.read_csv('Diagnosis_Data.csv') #Defining the features and target variablesX =

data.drop('Diagnosis', axis=1)

y = data['Diagnosis']
#Splitting the dataset into training and testing datasets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Defining the model

model = DecisionTreeClassifier(random_state=42)#Fitting the model

model.fit(X_train, y_train)#Making predictions

y_pred = model.predict(X_test)#Calculating the accuracy score

accuracy = accuracy_score(y_test, y_pred)#Printing the accuracy score

2) Using K-Mean Algorithm:
python codes for Automated Diagnosis Machine Learning Model using K-Means:from

sklearn.cluster import KMeans

from sklearn.metrics import silhouette_score

from sklearn.preprocessing import StandardScaler from sklearn.model_selection import

train_test_splitfrom sklearn.metrics import accuracy_score

from sklearn.metrics import confusion_matrix# Step 1: Load the data

data = pd.read_csv(r'diagnosis_data.csv')
# Step 2: Split the data into train and test setsX = data.drop('diagnosis', axis = 1)

y = data['diagnosis']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 42)#

Step 3: Standardize the data

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)X_test_scaled = scaler.transform(X_test)

Step 4: Fit the K-Means model kmeans = KMeans(n_clusters = 2)

kmeans.fit(X_train_scaled)

Step 5: Evaluate the model
Calculate the silhouette score silhouette_score(X_train_scaled, kmeans.labels_)

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345
Step 6: Make predictions y_pred = kmeans.predict(X_test_scaled)
Step 7: Evaluate the model# Calculate the accuracy

accuracy_score(y_test, y_pred) # Calculate the confusion matrix

confusion_matrix(y_test, y_pred) writing a clustering programe in python to separate the WBC and RBC Countimport numpy as np
from sklearn.cluster
import KMeans # Create Data
wbc_count = np.array([[4.5], [4.7], [4.8], [5.0], [5.2], [5.3], [5.5]])
rbc_count = np.array([[4.2], [4.4], [4.5], [4.7], [4.9], [5.0], [5.2]])
# Combine
Data data = np.concatenate((wbc_count, rbc_count), axis=1)# KMeans Clustering

kmeans = KMeans(n_clusters=2, random_state=0).fit(data)# Separating the clusters

wbc_cluster = kmeans.labels_[0:7] rbc_cluster = kmeans.labels_[7:14]# Print clusters

print("WBC Cluster: ", wbc_cluster)print("RBC Cluster: ", rbc_cluster)

4) Coding: Using Random Forest:
Import required librariesimport pandas as pd import numpy as np

from sklearn.model_selection import train_test_split from sklearn.ensemble import

RandomForestClassifier

from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
# Load data
data = pd.read_csv(r‘Diagnosis_data.csv')
# Split data into features and targetX = data.drop(['diagnosis'], axis=1)y = data['diagnosis']

Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Create random forest classifier object
clf = RandomForestClassifier(n_estimators=100, random_state=42)

Train the classifier clf.fit(X_train, y_train)

Make predictions on the test sety_pred = clf.predict(X_test)

Evaluate the model
print('Accuracy:', accuracy_score(y_test, y_pred)) print('Confusion Matrix:',

confusion_matrix(y_test, y_pred)) print('Classification Report:', classification_report(y_test,

y_pred))

Results:

```
Accuracy: 0.9649122807017544
Confusion Matrix: [[70  1]
[ 3 40]]
Classification Report: precision recall  f1-score support
```

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1337 - 1345

```
           B              0.96   0.99   0.97          71
           M              0.98   0.93   0.95          43
    accuracy                         0.96         114
   macro avg      0.97   0.96   0.96         114
weighted avg      0.97   0.96   0.96         114
```

5) Using SVM (Support Vector Machine):

```
# Import required librariesimport pandas as pd import numpy as np

from sklearn.model_selection import train_test_splitfrom sklearn.svm import SVC

from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

# Load data
data = pd.read_csv('leukemia_data.csv')

# Split data into features and target
X = data.drop(['Progression'], axis=1)y = data['Progression']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create SVM classifier object clf = SVC(kernel='linear', C=1)

# Train the SVM classifierclf.fit(X_train, y_train)

# Make predictions on the test sety_pred = clf.predict(X_test)

# Evaluate the model
print('Accuracy:', accuracy_score(y_test, y_pred)) print('Confusion Matrix:',

confusion_matrix(y_test, y_pred)) print('Classification Report:', classification_report(y_test,

y_pred))
```

6) Conclusion

In conclusion, our study aimed to predict leukemia progression using clinical and genomic features through various machine learning algorithms such as SVM, Random Forest, and Logistic Regression. After conducting a thorough analysis of the results, we found that the Random Forest model outperformed the other models in terms of accuracy, precision, recall, and F1 score. This model demonstrated superior predictive power in identifying patients at risk of disease progression, which can aid in personalized treatment planning and improve patient outcomes.

Our study highlights the potential of machine learning in leukemia research and the importance of incorporating clinical and genomic data for better disease management. The use of machine learning models can enable healthcare professionals to make data-driven decisions and provide personalized treatment plans for patients. However, it's important to note that these models should be used in conjunction with clinical expertise and not as a substitute for medical judgement.

Overall, our findings suggest that the Random Forest model can be a valuable tool in

predicting leukemia progression and can potentially contribute to better patient outcomes. Further studies are needed to validate these results and explore the potential of machine learning in leukemia research.

7) References:

[1] Machine Learning Model Development based on Brazil's Covid-19 Data Set by Dr.Kamel Alikhan Siddiqui, Dr.C V Narsimhulu / Dr. K Naggi Reddy. https://www.sumc.lt/index.php/se/article/view/1162, ISBN: 13925369, International JournalVol: Sept/2022, Vol2 no 43(2022) Pages: 403 to 412.

[2] Optimization, Clustering and SVM Analysis on Blood Cancer Cells – A Machine Learning Model of Neural Networks. ICASEM information Technology by 1Dr.Kamel Alikhan Siddiqui (Associate Prof, IT, LIET), 2Ms.Sumaya Begum (Assistant Prof, IT, LIET), 3Mrs.Khutaija (Assistant Prof, IT, Lords Institute of Engineering and Technoogy).

[3] **Incidence Networks for Geometric Deep Learning** by Marjan Albooyeh, Daniele Bertolini, Siamak Ravanbakhsh, https://arxiv.org/abs/1905.11460

[4] Rohokale, M. S., Dhabliya, D., Sathish, T., Vijayan, V., & Senthilkumar, N. (2021). A novel two-step co-precipitation approach of CuS/NiMn2O4 heterostructured nanocatalyst for enhanced visible light driven photocatalytic activity via efficient photo-induced charge separation properties. Physica B: Condensed Matter, 610 doi:10.1016/j.physb.2021.412902

[5] Signal and Image Processing: An International Journal(SIPIJ), ISSN:2229-3922

[6] Raghavendra, S., Dhabliya, D., Mondal, D., Omarov, B., Sankaran, K. S., Dhablia, A., . . . Shabaz, M. (2022). Development of intrusion detection system using machine learning for the analytics of internet of things enabled enterprises. IET Communications,doi:10.1049/cmu2.12530

[7] Comparison of GPU and FPGA Hardware Acceleration of Lane Detection Algorithm Full Text BY Mohamed Alshemi1, Sherif Saif2 and Mohamed Taher3, 1Ain Shams University, Egypt, 1STMicroelectronics, Cairo, Egypt, 2Electronics Research Institute, Egypt, 3Ain Shams University, Egypt