**Data Modernization and AI in health-care: A detailed study by Yash Shrivastava**

**Introduction: AI and Data Modernization**

Artificial Intelligence, Data Science, Machine Learning Ah! These words have been in the limelight in the entire last decade and it still fascinates techies, Businesses, and Students all around the world. In the last decade AI’s and its applications have sky rocketed, Researchers and Scientists have achieved some of the major breakthroughs in the technology by coming up with AI products that ranges from Automating tasks, making day to day life easier for people to solving bigger problems that Mankind was not able to engineer. Now AI is present in almost every domain we can think of from Oil/Gas, Health care, Manufacturing to Market Basket analysis, Recommendation systems used by different e-commerce and entertainment platforms.

Data Modernization refers to the process of supplanting the legacy database to modern cloud base systems that are faster, more secured and provides a lot more features for data security, transfer, Pipelines and speed.  
AI along with data modernization has given researchers the ability to efficiently create and deploy massive AI products like ChatGPT, Yolo, Bard etc..

Here in this blog we will focus on how AI and Data Modernization have revolutionize health care and clinical decision making.

**AI in Health care:**

AI has shown its mettle in almost every business and domain of work. Similarly, it has been a boon in the health care domain, From disease prediction to unearthing complicated patterns in blood, DNAs etc. AI has covered it all.

In this blog we will focus on how AI can help in :-

1. Early detection of serious health issues.
2. Using CNN/Vision models to comprehend medical imaging.
3. Using LLM models for auto OPD during the initial level of treatment.
4. Drug Development and Medical Research.
5. **Early detection of serious health issue:**

Everyday human body generates TBs of data, Doctors and medical researchers can record this data and use data science to exhume patterns that are unlikely to be uncovered by humans.

Machine learning can be used to predict serious health issues given we have trained our ML models with good historical data and relevant features. Similarly Neural Network models using convolution layers can be used to comprehend significant patterns in blood and other medical scans which can be used to predict serious health issues in the preliminary stages.

Data science can also be used to detect minute changes in different clinical signals of the body and can give warning signals to the user which can be a life saviour in certain emergency medical conditions. Smart wearable nowadays are a very good example of real time tracking of clinical signals and using AI to comprehend medical condition.

Since the health-care sector heavily relies on documentation, research, and proof-of-concepts, it is crucial to utilize data science methods that are straightforward and easy to explain and document. Typically, fundamental ML algorithms such as K-Nearest Neighbors (KNN), Decision Trees, Support Vector Machines (SVM), and Naive Bayes are preferred over more complex and less understandable options like Random Forest, other ensemble models, and Deep Learning. However, if we are building something very complex that is not possible with core machine learning algorithms, then we may need to use more complex algorithms. For this blog we used a data that was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals.The original study published the feature extraction methods for general voice disorders.

This data was used to determine if a person is suffering from Parkinson’s disease or not. We have following features in our data:

***name*** - ASCII subject name and recording number

****\*MDVP:Fo(Hz) \*****- Average vocal fundamental frequency

***MDVP:Fhi(Hz)*** - Maximum vocal fundamental frequency

***MDVP:Flo(Hz)*** - Minimum vocal fundamental frequency

***MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP*** - Several measures of variation in fundamental frequency

***MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA*** - Several measures of variation in amplitude

****\*NHR,HNR \*****- Two measures of ratio of noise to tonal components in the voice

****status**** - Health status of the subject (one) - Parkinson's, (zero) - healthy

***RPDE,D2*** - Two nonlinear dynamical complexity measures

***DFA*** - Signal fractal scaling exponent

***spread1,spread2,PPE*** - Three nonlinear measures of fundamental frequency variation

We trained our model using different ML algorithms and found KNN and RandomForestClassifier to be performing best on the data. Due to the reasons stated above we chose KNN to be the algorithm to be the best fit for this application.

The model was able to achieve an accuracy of 98%.

**ipynb file attached **

1. **Using CNN/Vision models to comprehend medical imaging:**

CNN is a type of neural network which is used when the data has spatial features for ex: images, matrix etc.. CNN consists of a convolution layer which performs convolution operation on each pixel of the image to generate a feature map and finally outputs of multiple convolution layers are combined to come up with a final feature map of the input image, This feature map is then used as the input to the fully connected feed forward network similar to that of a simple neural network. This property of CNN makes it a vital tool that can be widely used in the health care industry. Medical imaging analysis is the process of interpreting and extracting information from medical images, such as X-rays, MRI scans, CT scans, and ultrasound images. It has the potential to revolutionize the way diseases are diagnosed and monitored, but there are several complexities and challenges associated with this field.

Medical imaging analysis faces challenges such as image variability, costly data acquisition and annotation, data imbalance, inter- and intra-observer variability, complexity of anatomy and pathology, limited and noisy labels, and ethical and legal considerations. These challenges arise from differences in images, difficulties in obtaining and labeling data, imbalanced datasets, variability in expert interpretations, complex structures and subtle abnormalities, subjective labeling, and the need for ethical handling of patient data.

To tackle these complexities and challenges, it is essential to foster interdisciplinary collaborations among data scientists, radiologists, medical professionals, and domain experts. Utilizing advanced algorithms, implementing robust validation methodologies, employing data augmentation techniques, and establishing standardized imaging protocols and annotation guidelines can effectively address these challenges and enhance the accuracy and dependability of medical imaging analysis.

Accurate and early detection in medical diagnosis and treatment is crucial because it enhances patient outcomes, enables timely intervention, broadens treatment choices, facilitates disease monitoring, improves patient management, lowers healthcare expenses, and positively influences public health.

Today AI is at par or better than seasoned radiologists at beholding malignant tumors. Some patterns in blood and medical scans are difficult conundrums to comprehend for researchers. CNN models have automated this task and these patterns are easily identified by these models given enough labeled data is available.

These models can also be used to localize anomalies in images, highlighting the most important areas of interest. This can be used to track disease progression or the response to a treatment.This becomes very vital for early disease detection and research work.

For this blog we have trained a very basic CNN model with 3 convolution layers with kernel size 3X3 and 32, 64 and 108 filters respectively. We gave our model pictures of X-Rays of normal and pneumonia patients, With a very basic model and no data manipulation/Augmentation our CNN model was able to achieve 90%+ accuracy classifying the X-rays. With transfer learning and models trained on better and bigger data, it is very easy to achieve even higher accuracy.

In similar fashion CNN models can be used for medical imaging, Research and Disease prediction.

Google Colab link: <https://colab.research.google.com/drive/12oHjOBOQk0Kv3_TKgjA2GZ95QHKquSjW#scrollTo=81EVbyaf-elZ>

1. **Using LLM models for auto OPD during the initial level of treatment:**

NLP and LLMs have a wide range of applications in healthcare, such as clinical decision support, drug discovery, patient education, virtual assistants, chatbots, clinical documentation, risk assessment, and research analysis. These technologies leverage language processing capabilities to improve patient care, enhance decision-making, automate administrative tasks, and extract valuable insights from medical data. As NLP and LLMs continue to advance, their impact on healthcare is expected to grow, enabling more efficient and personalized healthcare delivery.

LLM stands for Large Language models which are basically AI models used for Natural Language Processing tasks, Some examples of LLMs are GPT, BARD, BERT etc.. LLM models are used to comprehend the language used by humans to perform AI tasks. LLMs uses bidirectional transformer that can take care of preceding as well as following text to understand the context, Apart from Bi-Directional transformers they also use attention mechanism to focus on the most relevant part of the text.

LLMs can play a crucial role in automating various aspects of the Outpatient Department (OPD) in healthcare. By leveraging NLP techniques, the process of patient appointment scheduling can be streamlined, enabling efficient handling of requests, rescheduling, and cancellations. NLP algorithms can analyze patient symptoms and medical history, aiding in triaging and directing patients to the appropriate departments or specialists. Additionally, NLP can support preliminary diagnosis by extracting pertinent information from patient-reported symptoms. Moreover, NLP-powered chatbots or virtual assistants can deliver patient education, address inquiries, and provide guidance on managing health conditions. The implementation of NLP in auto OPD systems can optimize resource utilization, enhance operational efficiency, and improve patient care.

Moreover, Combining speech and language models can be used to achieve direct verbal communication to achieve everything that has been discussed above.

1. **Drug Development and Medical Research**

Pharmaceutical industry has seen dramatic increase in digital data in the past few years. AI has the ability to analyse this huge volume of data effectively and coming up with discoveries that are preserved in the data. Drug design, chemical synthesis, drug screening are some of the important aspects of drug development where AI can be effectively used. It is believed that AI can reduce the cost and time taken for drug development by up to 70%. The discovery of any new drug molecule goes through 4 major stages 1. Drug Discovery 2. Pre-clinical research like animal testing 3. clinical research i.e. human testing 4. Post-market research.

The search for new targets, biomarkers, and compounds begins with the identification of potential therapeutic agents. To do this, researchers examine how various molecules, genes, and proteins interact with one another in order to determine which molecules have the greatest potential.

Astonishingly, Tire-I Pharma companies were able to develop COVID-19 vaccinations. This was made possible by their historically quick development and use of these vaccines. New drug development platforms that enable researchers to apply artificial intelligence and machine learning to model protein and cellular interactions in order to quickly advance the science have played a significant role in this innovation acceleration.

Researchers are no longer required to rely solely on conventional laboratory testing (and retesting). Researchers may employ AI to enable simulations on computers rather than testing in real-world conditions thanks to their improved grasp of the molecular and genetic structure of a patient and, for example, their tumour. To aid in determining the presence of high-potential candidates for treatment consideration and subsequent analysis, this system can process hundreds, even millions, of simulations.

AI-enabled drug discovery models can shorten the time it takes to develop a drug and get it to a patient faster by cutting months and years off the research process. Consider the work on AlphaFold2 by Google's DeepMind division, which uses AI to forecast the most efficient protein structures for novel medications, as just one illustration. Organizations in the health care and life sciences sector are already aware of AI's potential. The investments are now needed to take advantage of this quickly developing technology and support their current and future initiatives.

**Conclusion**

In above blog we saw the potential of AI in heath care and there is not doubt that AI and data modernization is going to change how decision making is followed in health care. All of this AI research is not possible without the rapid modernization of data. All major AI models requires large volumes of relevant data which is reliable. The use of cloud based databases has provided the flexibility to create such AI models, Also in the coming years we can say that we will more and more data to come up with AI products that will change the entire operations of health care and aid research and help to provide better treatment to patients. Advance AI like reinforcement learning can make a way for robot assisted surgeries, Robot assisted nursing etc.. It will be extremely exciting to see what AI is going to bring in the coming decade.