How to Cite:

Advancement of natural language programming, machine learning and electronic health records for the digital health science field

M. Srilakshmi
Assistant professor, Department of chemistry, V.r.siddhartha engineering college,Vijayawada-07
Corresponding author email: srilakshmi771977@gmail.com

Dr Nitin Purohit
Associate professor, Department of computer science, Kebri Dehar University, Ethiopia
Email: nitinpurohit111@gmail.com

Dr. P. Rajesh
Assistant Professor, PG Department of Computer Science, Government Arts College, Chidambaram - 608 102, (Deputed from Dept. of Computer and Information Science, Annamalai University, Annamalainagar) Tamil Nadu, India
Email: rajeshdatamining@gmail.com

Dr. D. Beulah David
Professor, Institute of Information Technology, Saveetha School of Engineering, Chennai
Email: beulahdavid.sse@saveetha.com

Dr. k. s. v. k. s. Madhavi Rani
Associate professor, Department of zoology, Ch.S.D.St. Theresa’s (A) College For Women, Gavaravaram, Eluru - 534003, Eluru District, Adhra Pradesh
Email: madhaviraniksvks@gmail.com

Dr. Harishchander Anandaram
Assistant Professor, Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India
Email: a_harishchander@cb.amrita.edu

Abstract---The widespread use of “electronic health record systems (EHRs)” in health care provides a large amount of real-world data, opening up new opportunities for medical trials. “Deep learning, a subset of machine learning (ML)”, has experienced a meteoric rise in popularity over the last six years, owing to advances in computing
power and the accessibility of enormous new datasets. “Natural language processing (NLP)” approaches have been used as an “artificial intelligence strategy” to obtain information from medical narratives in EHRs, as a huge quantity of useful clinical knowledge is contained in clinical stories. This NLP capacity may enable “automated chart review in clinical care to identify individuals with distinct clinical features” and decrease methodological heterogeneity in establishing “phenotypes, masking biological heterogeneity in allergy, asthma, and immunology research”. Aim of this research paper is to understand the advanced technologies such as “Machine Learning, Natural Language Programming” are helpful for digital health field. In this context, secondary data collection method is used to gather information related to this topic.

Keywords---machine learning, natural language programming, electronic health record systems, electronic, medical, health.

Introduction

“Electronic health records (EHRs)” systems have been more popular in US healthcare facilities over the last decade. As a result, the use of artificial intelligence (AI) (computer-assisted human tasks) in health treatment and practice is now becoming increasingly relevant in a variety of clinical settings. A few advanced AI studies in the area of respiratory illness and other care facilities, for instance, illustrate the practicability of interpreting "pulmonary function test (PFT)" outcomes and their corresponding diagnosis, timely identification of symptomless “left ventricular” disorder from a “12-lead EKG”, and early identification of “atypia” or carcinoma in situ” from a “breast tissue biopsy”. These efforts may merely scratch the surface of what “EHR-based informatics research” has to provide in the upcoming years.

Background of research
The "Clinical and Translational Science Awards (CTSA)" Scheme, the "Electronic Medical Records and Genomics (eMERGE)" Network, which connects DNA biorepositories with EHR systems, the "Clinical Data Research Networks (CDRN)" Program, which is assisted by the 'Patient-Centered Outcomes Research Institute (PCORI)" for relative evidence on the effect, and the "Observational Health Data Sciences and Informatics (OHDSI)" Program have all been developed to take advantage of this "Common Data Model (CDM)". Because a substantial quantity of patient data is contained in clinical narratives, the enormous potential of these major projects is heavily dependent on efficient and accurate mining methodologies resulting in extracting knowledge and improving clinical practice [1,2]. Clinical documentation such as the background of current disease, medical examination, "narrative description of physical examination" results, PFT report, and radiological or surgical reports, for instance, take up a large amount of time for physicians. An online health portal is now a significant source for gathering patient-reported outcomes (PRO) as well as a platform for interaction between physicians and patients mostly in free text. Unstructured data comprises more than 80% of present accessibility to healthcare information, according to specialists at "International Data Corporation". In order to obtain information from medical reports for biomedical studies, "natural language processing (NLP)" approaches have been used. Now that NLP systems have this power, “AI-assisted cohort” identification and/or matching from EHRs is possible, potentially making clinical studies or research virtual or, at the very least, simplified [2].

![Figure 2: Making predictions by using EHRs](Source: [2])

**Literature Review**

Deep learning is a prominent machine learning (ML) approach that is praised for eliminating the need for considerable human feature engineering, which is prevalent in previous approaches. “Deep learning (DL)” is defined as “neural network-based” or -inspired approaches that employ contemporary optimization algorithms and instructional strategies for the sake of this research. “Recurrent neural networks (RNNs)” are widely used to represent sequential information like
language, whereas "convolutional neural networks (CNNs)" are commonly used to model outputs like pictures [3]. Under the deep learning framework, researchers “include embeddings—dense, data-driven” vector illustration renderings of, say, a phrase. Most current neural networks use embeddings as their input layer; some extracted features are explicitly formed as part of bigger “neural networks”, whereas other embedding approaches lack the nonlinear effects that characterise neural networks [4].

There is evidence that DL for Medicinal NLP has become more commonly acknowledged, in addition to expanding in volume. This acceptability may be seen from the reality that deep learning techniques are increasingly being regarded as the systematic procedure, with no need to compare them to classic machine learning. DL-based NLP techniques have also spread beyond the computing field and into renowned clinical publications, despite their origins in the CS and NLP areas [5]. The use of CNN-based algorithms in image-level diagnosis has been highly fruitful. This is partly because CNNs have surpassed individual intelligence in “object-classification” activities, in which a CNN develops to categorise the item included in an image. These very same infrastructures have shown stellar performance in learning algorithms, in which a CNN is fine-tuned on a much dataset consisting associated with the task of interest after being trained on a massive dataset un-associated with the assignment of involvement “(e.g., ImageNet, a dataset of millions of widely accepted everyday items) (e.g., medical images)”. The system learns the natural characteristics of images—straight lines, curves, colourations, and so on—in the first stage, and then the higher-level levels of the system are retrained to discriminate between diagnostic situations in the second phase [6].

Figure 3: Neural network layers and deep learning concept in digital health science
(Source: [7])
“Object identification and segmentation techniques”, on the other hand, define key areas of a picture that belong to certain items. Picture data is fed into CNN algorithms, which iteratively twist it via a “series of convolutional and nonlinear processes” until the actual data matrix is converted into a probability distribution across all possible image classes “(e.g., medical diagnostic cases)”. “Deep-learning models” have impressively achieved physician-level accuracy in a wide range of diagnostic tasks, including distinguishing moles from "melanomas", "diabetic retinopathy", cardiovascular risk", and consultations from fundus and "optical coherence tomography (OCT)" images of the eye, “breast lesion identification in mammograms", and spinal analyzation with “magnetic resonance imaging (MRI)” [8]. Even across clinical disciplines, a single “deep-learning model” has been shown to be excellent in diagnosis (e.g., “radiology and ophthalmology”). As DL technology becomes more recognised and popular, informaticians and physicians will be more inclined to use it in clinical settings. Because this is both a potential concern, health practitioners must embrace and assess the risks correctly. In our data study, researchers were able to confirm a number of frequently held beliefs. Since early, effective CNN-based approaches, CNNs have conquered the Text Classification issue [9]. LSTMs with NER showed a similar impact “(typically cast as a sequence labelling problem)”. Recent innovations of “deep learning (DL)” for professional NLP published in NLP settings rather than in computing or clinical forums, as predicted; there is a lag in the acceptance of cutting-edge methods in the information systems field and an even greater lag in the medical world [10].

Other assessments, on the other hand, were unexpected when contrasted with popular opinion. Whereas prior clinical NLP systems appeared to make considerable use of knowledge-based resources, only “17.9% of DL-based techniques” did. Whereas prior research indicated “French” to be the most prevalent “non-English language in clinical NLP”, the findings reveal a previously unknown truth regarding non-English clinical NLP: Chinese has higher representations than French when it comes to deep learning. While some have believed that preprint servers include all cutting-edge DL research, researchers discovered that “only 16.5% of the articles in the sample were released as preprints prior to publication. Projections this review” has identified several possible future DL developments in “clinical NLP”. Domain adaption and
transferable learning algorithms are critical due to the high expense of labelling medical databases as well as the confidentiality problems associated with sharing in-domain training data. Furthermore, due to a lack of compelling results, there has been a very little detailed examination of this challenge from a deep learning approach. With the advent of effective pre-trained algorithms like BERT, researchers anticipate a rapid increase in the application and refining of learning algorithms. Regardless, researchers feel that medical expertise resources are underused. Deep learning’s core belief has been to "let the algorithms decide what’s essential," rather than handcrafting features, DL structures and sources still require human involvement, as illustrated by a recent movement to address inductive bias “(eg, gender biases found in word embeddings)”. “Knowledge resources” have the potential to give a calculable and objective way of guiding data-driven DL techniques, and the healthcare field is exceptionally well-equipped with such resources. Deep learning is often not effective, as evidenced by other subspecialties employing ML for therapeutic purposes. This means that it frequently fails to outperform simple models like logistic regression.

**Methodology**

Research methodology is a process of collecting data by which researchers can gather relevant data related to research topic. Secondary data collection method is used for this research paper to perform this. Different databases such as Google scholar, PubMed, ProQuest are used to collect articles and journals related to technologies such as ML, Deep Learning and electronic health record system for managing digital health system. Large volumes of comprehensive longitudinal patient data, such as clinical history, lab tests, prescriptions, therapies, and prognoses, are gathered, maintained, and made accessible virtually. These large medical datasets are important sources of information for basic and preclinical research.

**Analysis and Discussion**

![Figure 5: “Comparisons between deep learning techniques and traditional machine learning, over time”](Source: [13])
There were no in-depth assessments of the limits or shortcomings of DL approaches for medical NLP in the current batch of literature used in this research. There was no research into the fundamental causes of the failure in the few situations \((n = 12)\) where “DL failed to outperform standard ML”. Research on the comparative pros and drawbacks of DL based on the data quantity, data integrity, language, clinical specialism, bioinformatics activity, and, more importantly, the incorporation of knowledge-based resources is all missing. As a result, researchers believe that scientific work on the limits of DL approaches for clinical NLP tasks is crucial. For instance, DL technique efficiency advancements frequently rely on significant computing resources that use a lot of energy, which is bad for the environment. “Reinforcement learning (RL)” is a set of approaches for teaching computer agents to engage effectively with their surroundings, usually to accomplish specified goals. This learning might take the form of experimentation, presentation, or a combination of the two. A continuous feedback mechanism of reward and punishment trains an agent to better fulfil the tasks at hand as it performs actions in its surroundings. Learning from expert demonstrations can take two forms: learning to anticipate the expert’s behaviours directly through “supervised learning (i.e., imitation learning)” or implying the professional’s goal through inference (i.e., inverse RL). It is necessary to have an establishing guideline that can receive sensory information about the environment as outputs and input the next behaviours for the agent to execute in order to properly train an agent. “Deep RL”, which uses a “deep-learning model” as the sample parameter, appears to be promising. Robotic-assisted surgery is one area of medicine that can improve from deep RL (RAS). Nowadays, RAS relies heavily on a surgeon teleoperated guiding a robot's tools. By employing “computer vision models” (e.g., CNN's) to assess surgical surroundings and RL approaches to learn from a surgeon’s physical actions, deep learning can improve the resilience and flexibility of RAS. Suturing and knot tying, for example, are incredibly tedious and time-sensitive medical operations that might benefit from mechanization and efficiency. “Computer vision techniques” (“e.g., CNN’s for object detection/segmentation and criteria”) can, for example, re-create the scenery of an open wound from visual information, and a “suturing or knot-tying trajectory” can be derived by attempting to solve a dynamic optimization issue that tries to determine the shortest path while properly accounted for external constraints like combined boundaries and obstacles. “Image-trained RNNs” may also learn to tie knots on their own by “learning sequences” of events from surgeons, in this instance physical motions. These methods are very useful in completely “autonomous robotic surgery” and minimally invasive surgery. Considering “modern laparoscopic surgery (MLS)”, which involves making numerous small cuts in the body to implant a variety of equipment, such as cameras and surgical instruments, which surgeons then teleoperate. “Deep imitation learning, RNNs, and trajectory” transfer techniques may completely automate specific surgical procedure “teleoperated manipulation” jobs. Automating repetitious activities is more time-critical in MLS than it is in surgical intervention. For example, rather than a few seconds “in open surgery, it may take 3 minutes” in MLS to make a knot. Accurately localising an object’s position and orientation in the region of surgical scenes is one of the most difficult issues in semiautonomous teleoperation. “Recent pixel-wise instrument segmentation” algorithms based on an upgraded “U-Net architecture CNN” are beginning to show promise in this
Data collecting is another obstacle to the advancement of “deep learning in surgical robotics”.

**Conclusion**

“Deep learning” is adaptive to areas where input data is complicated and requires specific treatment, in addition to “Computer Vision (CV)”, NLP, and RL applications. For the sake of illustration, researchers will look at genomics as an illustration of how DL has been extended beyond traditional “(e.g., CNN- or RNN-based)” algorithms to operate with “nonimage, nontemporal” information formats. “Modern genomic techniques” gather a vast range of data, from a person’s “DNA sequence” to the number of protein molecules in their blood. Deep learning has a lot of potential for improving the methods for analysing these metrics, which will help physicians produce more accurate therapies and diagnostics. In genomics, a standard pathway for developing a deep-learning platform is receiving raw data transforming it into input “data convolution layers”, and putting these transform “through neural networks” that eventually power specialised biological applications.

One set of possibilities is "genome-wide association (GWA)" studies, which are huge case-control studies aimed at identifying causative genetic changes that impact certain behaviours. GWA investigations need methods that scale to very large patient groups and account for hidden variables. These problems may be solved using "deep learning optimisation" tools and techniques, such as “stochastic optimization” and other current algorithms paired with software solutions for parallel computing, as well as simulation strategies that account for unknown factors. Models that incorporate external methodologies and available sources of biomedical information into “GWA studies”—for example, imaging techniques or measurement techniques of splicing and other intermediate “molecular phenotypes”—might benefit from machine learning in the coming years to more correctly determine disease-associated causal genetic variations. Clinicians can propose therapies and give more accurate diagnoses if they understand the genetics of illness. Identifying whether new variations in a patient's genome are medically indicated is a major difficulty for doctors. Since of bigger, more proximate indicators and more advanced training information, intermediary biochemical configurations may be easier to quantify than human attributes. These two characteristics make the challenge well-suited for machine learning, which has already demonstrated effectiveness in estimating splicing and activator protein interaction. Genomic data can potentially be used as a direct biomarker for illness development and progression. Blood, for example, includes microscopic pieces of cell-free DNA generated by cells throughout the body. Organ rejections (i.e., the “immune system” rejecting transplant cells), severe infection, and early-stage malignancy are all non-invasive signs of these fragments. Predictive diagnostics using cell-free DNA is a great achievement: foetal DNA found in the mother’s blood shows “chromosomal abnormalities” and can disclose the whole sequence of the foetus. “Deep-learning systems” can improve the performance of “biomarker assays targeting” “DNA sequences”, “methylation”, “gene expression”, “chromatin profiles”, and many other measurement techniques. Biomarker information is “often noisy and requires complicated analysis” (e.g., “to decide exactly whether cell-free DNA is reflective of cancer”).
References


