

Detecting Clinical Errors in Electronic Health Records Using Machine Learning

¹Sahana K K(4MC23CS141), CSE, Malnad College of Engineering (Hassan).

²Poorvika H J (4MC23CS115), CSE, Malnad College of Engineering (Hassan).

³Sandesh S(4MC23CS143), CSE, Malnad College of Engineering (Hassan).

⁴Rizwan M(4MC23CS138), CSE, Malnad College of Engineering (Hassan).

⁵Dr. H.M Keerthi Kumar, Associate Professor, Malnad College of Engineering (Hassan).

Abstract— Electronic Health Records or EHRs are now widely used in healthcare to store information digitally. These records often have mistakes like duplicate entries, wrong prescriptions and missing data. These issues can hurt safety and the decisions doctors make.

This paper looks at how techniques like Machine Learning, Deep Learning, anomaly detection and Natural Language Processing can help find and fix errors in EHR systems. It also talks about how to prepare data and different healthcare datasets and models used to detect errors.

The study finds that smart systems can make data more accurate improve safety and help to make better decisions with EHRs. This are very important, for healthcare.

The study also shows how it can be improved with these systems.

I. INTRODUCTION

Electronic Health Records are really important in handling information in healthcare. Hospitals use them to keep track of things, like diagnoses, prescriptions, lab results and treatment history.

Systems make things more efficient. They can have mistake. Some common problems are entries, duplicate records, missing information and prescription errors.

In the past people checked for these errors by hand. Its a timewasting process that doesn't work well with big datasets. As healthcare data grows, we need automated ways to deal with it.

The Electronic Health Records include things like diagnosis reports and prescriptions and lab results and doctors' notes and patient monitoring information and discharge summaries and treatment history and radiology reports and insurance details and patient demographic information.

Using Health related Records makes healthcare a lot easier for everyone involved. These problems can be things like wrong entries or abnormal prescriptions or inconsistent diagnoses or missing information or medication errors or abnormal healthcare behavior.

So It must be accurate and reliable all the time. This helps them provide healthcare to their patients. Electronic Health Records are really crucial for healthcare. Healthcare organizations rely on Electronic Health Records to work well and to provide the healthcare possible. These records help them make decisions about patient care and things, like that. Mistakes in Health Records are really bad for healthcare and patient safety. If the information about a patient's health is wrong it can cause a lot of problems. The old way of checking healthcare information is mostly done by people who look at it manually and use rules to validate it. This way is not very good because the amount of healthcare information is always growing and it is very complex. There is too much information for people to check.

New healthcare systems that use Learning Machine and AI are better. It can automatically find patterns in healthcare information suspicious behaviour, wrong prescriptions, duplicate records and inconsistencies in diagnoses. This are very useful for Health Records. This helps doctors to find problems that they might not see otherwise.

This paper is a review of how Machine Learning's used to find mistakes. It talks about how to get healthcare information ready for use how to find anomalies how to monitor healthcare the challenges of keeping healthcare information safe and what the future of healthcare systems will look like.

II. BACKGROUND AND THEORETICAL STUDY

a. Electronic Health Records

The systems help doctors and nurses store information. They help manage information.

This information includes things, like reports, prescriptions and the patient history of the patient.

Electronic Health Record systems make it easy for healthcare professionals to access Electronic Health Record data and share the Electronic Health Record data.

This type of systems also works with healthcare services like cloud services and telemedicine and healthcare analytics and smart healthcare monitoring platforms.

b.Clinical Errors in Healthcare

Sometimes mistakes happen in healthcare because of information or duplicate records or medication mistakes or missing information or bad laboratory results or wrong diagnosis reports or communication problems.

Clinical errors can really affect the quality of healthcare. Keep patients safe. If we do not find these mistakes away, they can cause big problems and cost a lot of money to fix.

Some common mistakes in healthcare are:

1. Wrong diagnosis
2. Duplicate Electronic Health Records
3. Medication mistakes
4. Prescription errors
5. Missing laboratory results
6. Wrong patient information
7. Strange patient behavior
8. Wrong discharge papers
9. Wrong laboratory reports
10. Mistakes, in healthcare paperwork

c.Machine Learning in Healthcare

Machine Learning is really useful in healthcare because it can automatically look at healthcare data and find patterns that're not normal. Machine Learning systems in healthcare can look at healthcare records. Say if they are okay or not okay by checking the details of the healthcare records. There are a ways that Machine Learning systems in healthcare work including supervised learning and unsupervised learning and semi supervised learning.

Machine Learning in healthcare uses things like Random Forest and Decision Tree and Logistic Regression and Support Vector Machine and Naive Bayes and Isolation Forest and Gradient Boosting to find things that're not normal in healthcare.

d.Deep Learning in Healthcare

Deep Learning is also really useful in healthcare because it can automatically learn about patterns in healthcare data. Deep Learning systems in healthcare are good at finding things that're not normal and understanding what healthcare workers write. Deep Learning systems in healthcare use things like Recurrent Neural Networks and Long Short-Term Memory networks and Autoencoders and Convolutional Neural Networks and Transformer models to make healthcare better. Deep Learning

in healthcare is really good at looking at sets of healthcare data and understanding things that are not organized.

e.Natural Language Processing in Healthcare

Natural Language Processing is used in healthcare to look at things that are not organized like what doctors write down.

Natural Language Processing in healthcare looks at things like what doctors say. What patients say and medical reports and xray reports.

Natural Language Processing in healthcare makes it better to understand what people write and find things that're not normal.

Natural Language Processing in healthcare uses things, like tokenization and stemming and lemmatization and Named Entity Recognition and text embedding and sentiment analysis to make healthcare better

III. LITERATURE REVIEW

[1]Machine Learning Based Healthcare Systems Learning techniques work well in detecting mistakes in healthcare data.

This models are good, at finding patterns. Deep Learning helps make it more accurate when dealing with complicated datasets. People also use Natural Language Processing to go through text and get useful information out of it. Lately researchers have been looking into Explainable AI. This helps doctors see how predictions are made. It makes them trust these systems more. Machine Learning and Deep Learning are really helpful here.

[2]Deep Learning Based Healthcare Systems

By using Deep Learning models, we can understand healthcare data and make predictions about what will happen.

This type of techniques are very important for making healthcare systems better.

They help us look at healthcare data. Identify things and make good decisions.

They help us look at healthcare data. Identify things and make good decisions.

This systems are especially useful because they can learn on their own to find relationships in healthcare data.

This ability makes Deep Learning systems very valuable for understanding healthcare data.

These learning techniques are important because systems can learn on their own.

This type techniques are very good at understanding patterns, in healthcare data

[3] NLP Based Healthcare Systems

Natural Language Processing is a help for doctors in healthcare. It looks at what doctors write like notes and reports. It breaks down the text into parts finds the words and figures out what they mean. This helps doctors understand what patients need. Natural Language Processing is really useful for doctors it makes their job a lot easier. Doctors use to see if a patient is having a reaction to medicine. It helps them understand reports and make sense of them. It looks at doctors notes and reports to find information for patients. This way doctors can give patients the care they need. These techniques help doctors understand what patients need. Doctors use this technique to understand reports and make decisions.

[4] Explainable Artificial Intelligence in Healthcare

Explainable Artificial Intelligence in Healthcare makes Healthcare more transparent and easier to understand. This is done by finding the Healthcare features that are used to make Healthcare predictions and detect anomalies in healthcare. Some techniques like SHAP and LIME. Feature importance analysis helps make Healthcare monitoring more transparent and trustworthy. Explainable Artificial Intelligence in Healthcare also makes Healthcare more reliable and acceptable to doctors and clinicians who use Healthcare systems.

[5] Anomaly Detection Techniques in Healthcare

There are techniques used to detect anomalies in healthcare, such as Isolation Forest and Autoencoders and Clustering Based Detection and Statistical Analysis and Local Outlier Factor and Tensor based analysis. These techniques help find behavior, in healthcare and suspicious patterns in healthcare. They make Healthcare monitoring systems better and improve real-time Healthcare analytics.

IV. HEALTHCARE DATA PREPROCESSING

Healthcare datasets collected from Health Records systems often contain duplicated records, incomplete healthcare information, missing healthcare values, abnormal prescriptions, inconsistent laboratory reports, and diagnosis inconsistencies. Preprocessing plays an important role in healthcare analytics because accurate healthcare preprocessing improves healthcare classification performance and healthcare anomaly detection accuracy.

Important preprocessing techniques utilized in healthcare analytics include:

1. Data cleaning
2. Duplicate removal
3. Missing value handling
4. Data normalization
5. Feature encoding
6. Feature selection
7. Noise removal
8. Healthcare text preprocessing
9. Dimensionality reduction
10. Healthcare data balancing

Structured healthcare information including laboratory values, diagnosis codes, medication details, radiology reports, and prescriptions are converted into numerical healthcare features suitable for analysis.

NLP techniques are utilized for preprocessing physician notes, diagnosis summaries, and discharge reports.

Data preprocessing significantly improves healthcare analytics performance, healthcare consistency, and intelligent healthcare monitoring efficiency.



Fig. 2: Healthcare Data Preprocessing Workflow

V. CHALLENGES IN CLINICAL ERROR DETECTION

Clinical error detection in Health Records is a challenging task because healthcare datasets are highly complex, continuously evolving, and contain large volumes of structured and unstructured healthcare information. Important healthcare challenges include:

1. Incomplete healthcare information
2. Missing healthcare values
3. Healthcare data imbalance
4. Unstructured physician notes
5. Abnormal healthcare behavior
6. High dimensional healthcare datasets
7. Real-time healthcare monitoring complexity
8. Healthcare privacy and security issues
9. Limited healthcare interoperability
10. Large computational requirements

NLP required for analysing physician observations, diagnosis summaries, and discharge reports because such healthcare information is usually unstructured.

Deep Learning based healthcare systems improve healthcare anomaly detection performance but require large healthcare datasets and high computational resources. Healthcare data privacy regulations and healthcare confidentiality requirements also increase the complexity of healthcare analytics systems.

VI. WORKFLOW OF CLINICAL ERROR DETECTION SYSTEM

The workflow of the proposed healthcare monitoring framework consists of healthcare data collection, preprocessing, healthcare feature extraction, Machine Learning analysis, anomaly detection, healthcare validation, and healthcare alert generation.

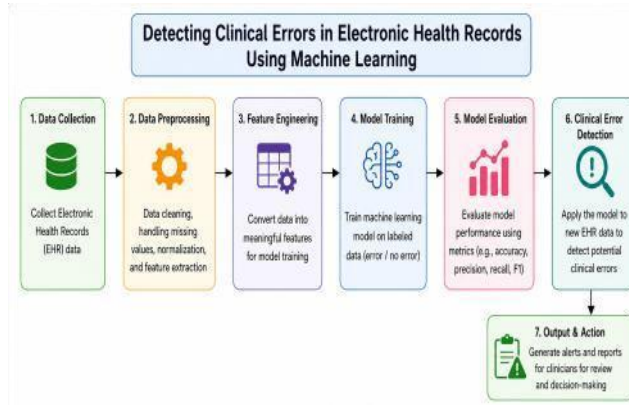


Fig. 3: Workflow of Clinical Error Detection System

VII. PERFORMANCE METRICS

Healthcare anomaly detection systems are evaluated using multiple healthcare performance metrics.

A. Accuracy

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Accuracy measures the overall healthcare classification healthcare anomaly detection system.

B. Precision

$$= \frac{TP}{TP + FP}$$

Precision measures the percentage of correctly identified healthcare anomalies among all predicted anomalies.

C. Recall

$$= \frac{TP}{TP + FN}$$

Recall measures the capability of the healthcare system to correctly identify actual healthcare anomalies.

D. F1-Score

$$2 \times \text{Precision} \times \text{Recall}$$

$$\text{F1-Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score provides balanced healthcare performance evaluation

TABLE I: Performance Comparison

Model	Accuracy	Precision	Recall
Decision Tree	91%	89%	88%
Random Forest	96%	95%	94%
Support Vector Machine	93%	92%	91%
Neural Network	98%	97%	97%

VIII. FUTURE RESEARCH DIRECTIONS

Healthcare systems of the future are going to use Artificial Intelligence to make healthcare monitoring better and to find problems in healthcare more accurately.

Artificial Intelligence is going to be very important in this.

Explainable Artificial Intelligence techniques will make healthcare more transparent and easier to understand. This means healthcare professionals will be able to understand what the smart healthcare systems are predicting and why they are finding problems.

Healthcare systems that use Blockchain technology will keep healthcare information secret make sure it is correct and protect healthcare privacy.

Federated learning approaches will help healthcare organizations work together without sharing healthcare information.

This way they can work together without putting information at risk. Cloud based healthcare systems and healthcare devices

that use the Internet of Things will make real time healthcare analytics and healthcare monitoring better.

New Deep Learning architectures, like the ones that use Transformer technology will make healthcare sequence analysis, healthcare prediction systems and healthcare text interpretation more accurate.

Future healthcare systems will also use devices, systems that monitor patients from far away smart healthcare assistants and automated systems that help make healthcare decisions.

These future healthcare systems will be very smart. Use Artificial Intelligence to make healthcare better.

IX. CONCLUSION

Electronic Health Records are a part of modern healthcare systems.

Healthcare organizations use healthcare platforms to do things. They. Monitor healthcare data. They also analyze this data. Talk about healthcare.

Healthcare organizations use healthcare platforms to do things. They. Monitor healthcare data. They also analyze this data. Talk about healthcare.

The problem is, healthcare datasets often have errors. These errors include information. There are also prescriptions. Sometimes there is duplicate or missing information.

The goal was to improve Electronic Health Records. This survey paper looked at techniques like Machine Learning, Deep Learning finding anomalies and Natural Language Processing. These techniques are used to find mistakes in Electronic Health Records.

The healthcare monitoring systems we looked at showed that using Machine Learning for healthcare analytics makes a difference. It helps find anomalies makes healthcare information more consistent. Improves the efficiency of healthcare monitoring. It also makes healthcare more transparent. Helps with decision making.

Using Machine Learning in healthcare systems also makes patients safer helps healthcare organizations work better and improves the way healthcare is monitored.

In the future healthcare systems will use technologies, like Explainable Artificial Intelligence, cloud-based healthcare analytics, federated learning, blockchain to secure healthcare information IoT to monitor healthcare and intelligent Deep Learning frameworks. These technologies will make healthcare monitoring systems smarter and improve patient

safety. Electronic Health Records and Machine Learning will keep playing a role in this.

REFERENCES

- [1] K. Noor et al., "Detecting Clinical Intent in Electronic Healthcare Records Using Machine Learning," IEEE International Conference on Healthcare Informatics, 2024.
- [2] G. Dei Mazar and A. Sheikh Taheri, "Machine Learning Models to Detect and Predict Patient Safety Events Using Electronic Health Records," 2023.
- [3] M. Li et al., "Identifying Hospital Patient Safety Problems Using Ensemble Machine Learning Models," International Journal of Clinical & Medical Informatics, 2018.
- [4] H. Niu et al., "Detecting Anomalous Sequences in Electronic Health Records Using Higher-Order Tensor Networks," Journal of Biomedical Informatics, vol. 135, 2022.
- [5] L. Francis et al., "Anomaly Detection in Electronic Health Records Using Machine Learning Techniques," IEEE SISIMPACT, 2025.