
Predicting Disease Outbreaks Using Healthcare Data Trends

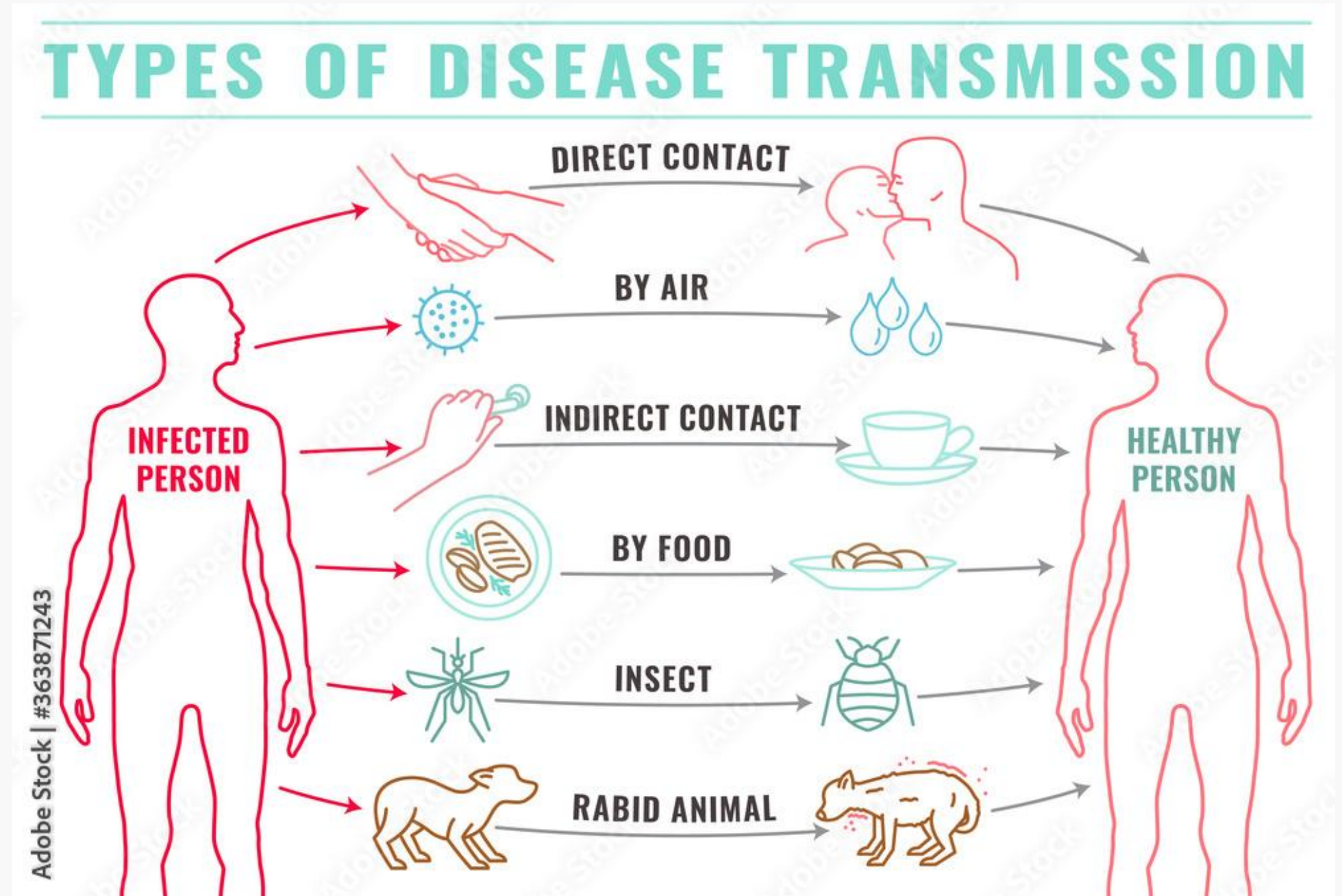
Presented by: [SNEHA BABARIYA]

[STUDENT NUMBER: 2409818]

[Research Methods and Professional Development - ACFA7004]

Definition of Outbreak in NHS

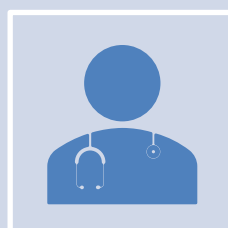
- an incident in which at least 2 or more people affected by the same infectious disease are linked by time, place, or common exposure



Research Context/Background



Global Health Concerns: These rising infectious diseases are global health concerns since they threaten all people's health and need immediate diagnosis systems to contain their harmful potential.



Role of Data Science: Data science has been changing the process of disease prediction by allowing the processing of such a large quantity of healthcare data via the usage of Artificial Intelligence (AI) and Machine Learning (ML) models to search for early warning carriers [1].



Need for Proactive Approaches: ML-driven analytics, when done proactively, helps in disease intervention at a stage before it becomes a nuisance, thereby reducing healthcare responsibility worldwide and helping curb disease transmission rates.

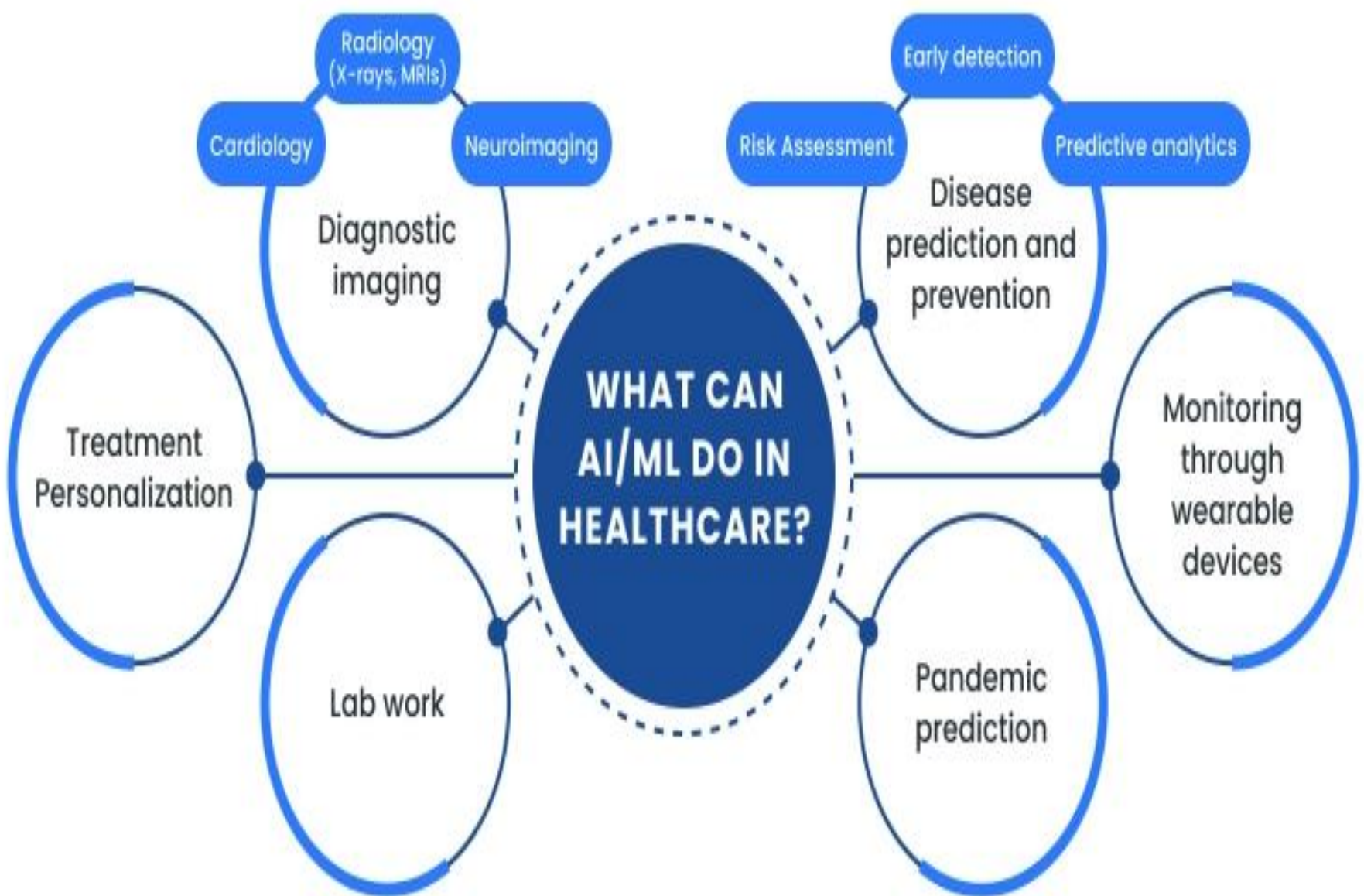


Figure: AI in Health-Care

Research Problem



1

Existing Challenges: Traditional disease surveillance has significant delays due to manual reporting and limitations in responding to outbreaks as they occur.

2

Data Utilization Gap: Existing systems do not utilize these data sources for predictive purposes and in real-time, such as with IoT, electronic health records, and social media [2].

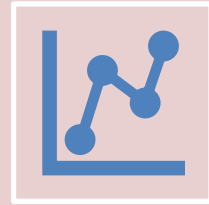
3

Unresolved Issues: Existing predictive models are largely inaccurate due to a lack of data quality consistency, bias, and poor feature selection, which makes it an unresolved issue to reliably detect early outbreaks.

Research Aim and Objectives

Research Aim:

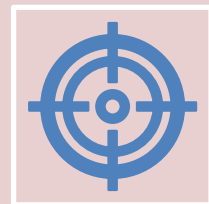
The aim is to develop a machine learning model in Python to be used to predict and analyze healthcare data trends as well as the likelihood of possible disease outbreaks occurring.



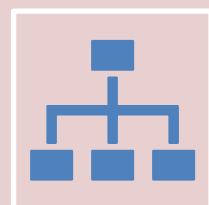
Data-Driven Insights: Collecting and preprocessing healthcare data, including patient records, social media trends, and environmental factors [1].



Model Development: To perform on implementing and comparing the developed machine learning and deep learning models for outbreak prediction



Performance Evaluation: Assessing model accuracy using key metrics like precision, recall, and F1-score to ensure reliability



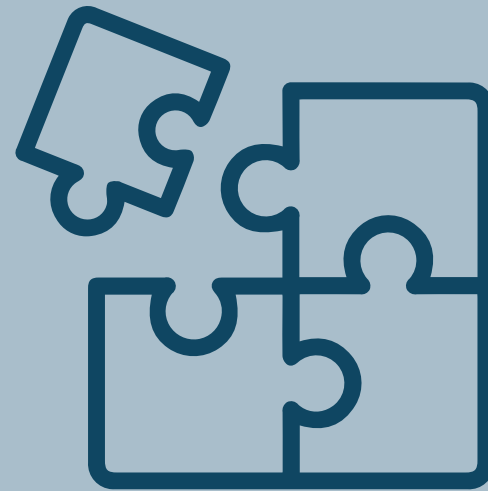
Comprehensive analysis: To discuss and compare the model's performance with traditional processes

Literature Review and Gap Identification



Existing Studies

- Several existing studies use ML-based models to make disease outbreak predictions, which have indicated significant potential but they face issues with execution in real-time with great accuracy and reliability.



Identified Gaps

- Current studies are lacking in the comprehensive integration of multiple healthcare data resources including EHR, IoT, and social media trends [3].



Research Focus

- The goals of this study are to increase the accuracy of prediction by using deep learning techniques to improve outbreak detection.

Research Methodology

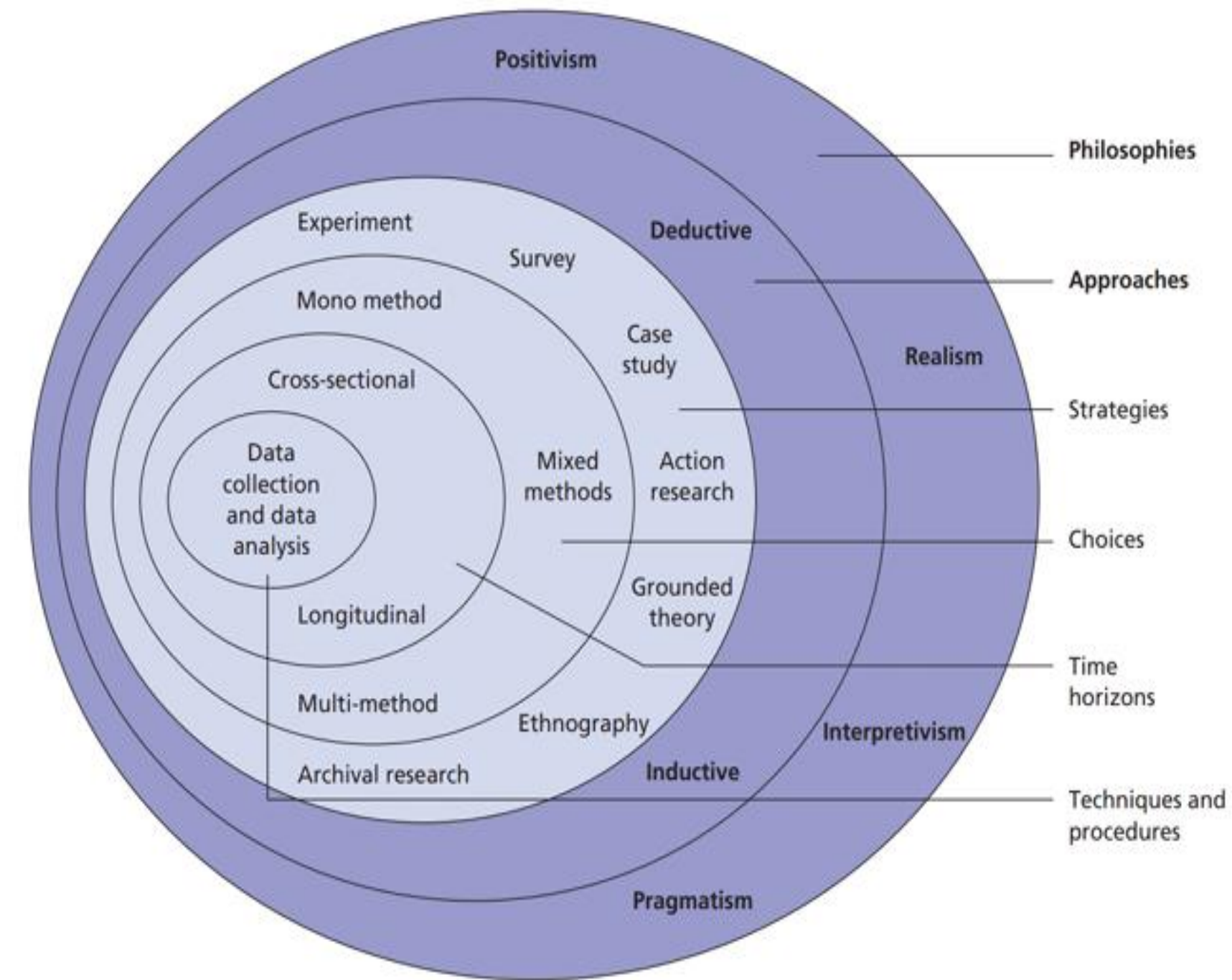


Figure: Figure: Research Onion Model

1. Research Philosophy

- **Positivism:** The study will follow a structured approach, using objective data for disease outbreak prediction.

2. Research Approach

- **Deductive:** Hypotheses will be tested using machine learning models on healthcare datasets.

3. Research Strategies

- **Experiment:** Evaluating ML models for accuracy in outbreak prediction.
- **Case Study:** analyzing real-world disease trends for validation.

4. Time Horizon

- **Cross-Sectional:** Data will be analyzed at a given point to identify patterns.

5. Data Collection & Analysis

- **Quantitative Approach:** Structured datasets, ML model evaluation, statistical validation (precision, recall, F1-score).

Methodology (Contd...)

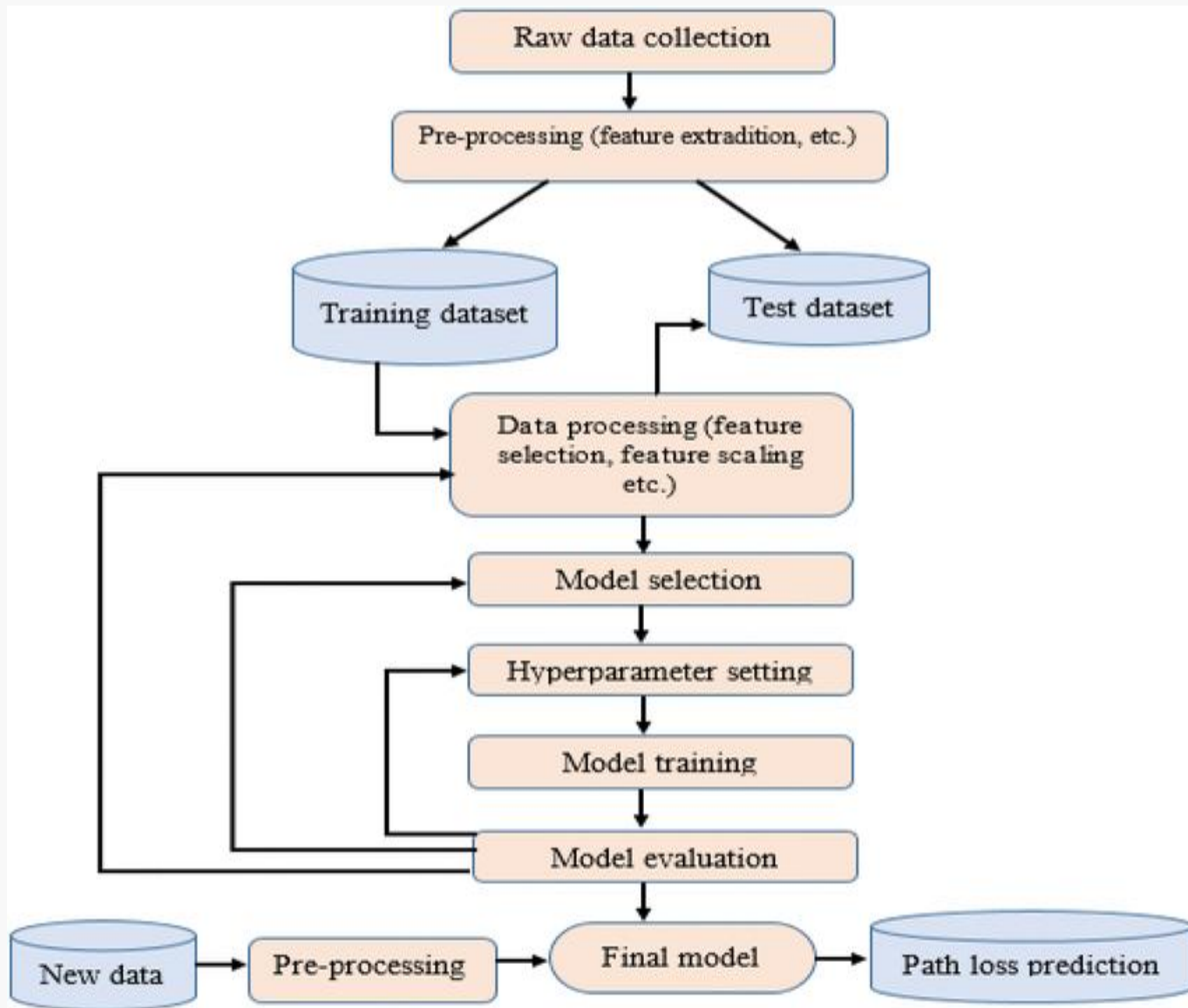


Figure: Figure: Flow chart for Model developemnt

Steps

Data Collection: To Gather disease outbreak datasets from Kaggle, electronic health records.

For now still dataset are not final I am searching on best dataset. Then I will work on model selection.

Model Selection: The predictive model will include ML Tool & techniques such as in python Random Forest and SVM alongside the deep learning methods LSTM and CNN for pattern detection [4].

Methodology (Contd...)

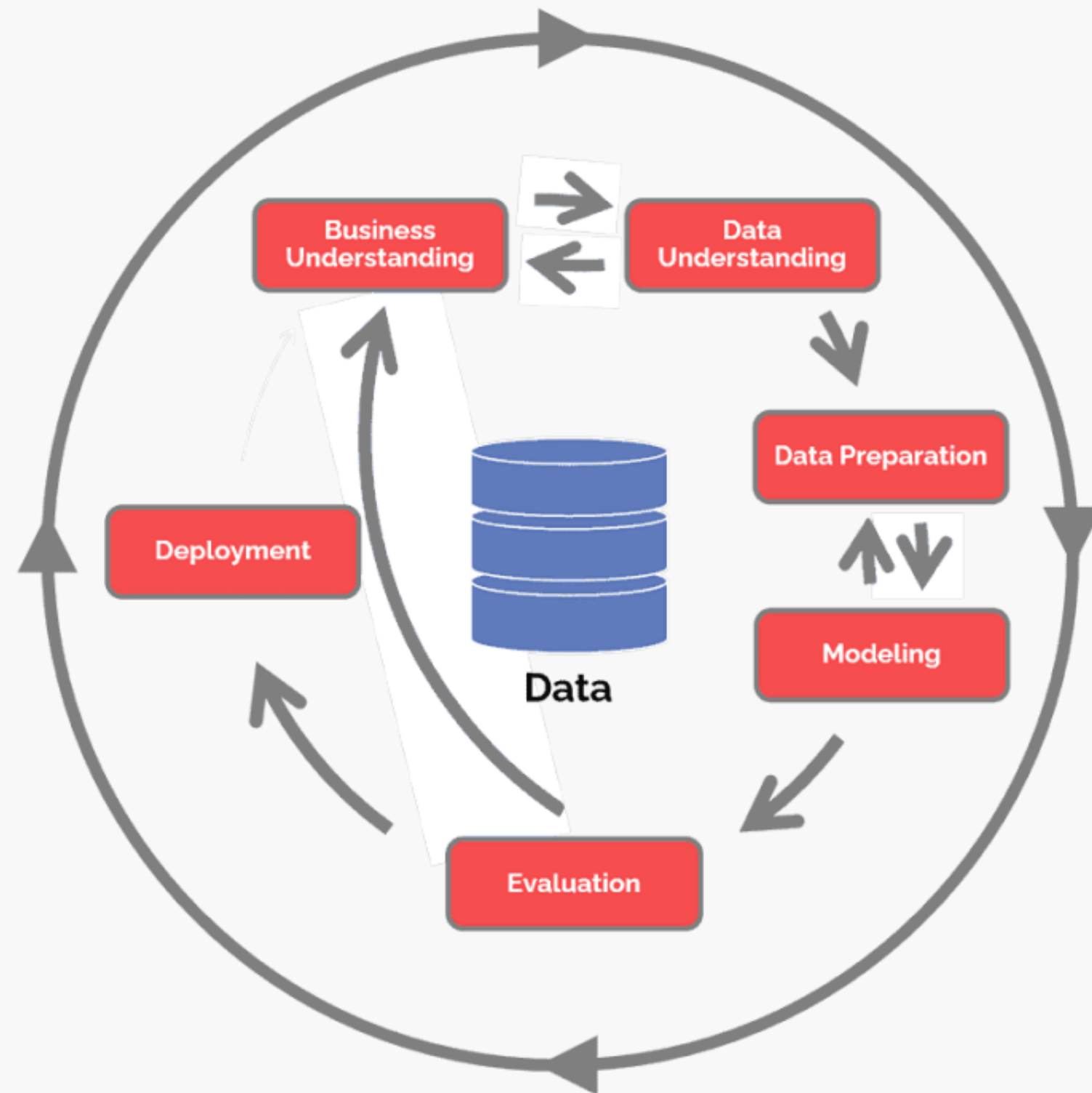


Figure: Experimental Work by following Crisp-DM model

I select CNN and LSTM because this method used in COVID-19 To predict the number of COVID-19 cases in Brazil, India, and Russia, three deep learning models, including CNN, LSTM, and the CNN-LSTM have been developed . The accurate analysis confirmed that the LSTM model had the highest performance.

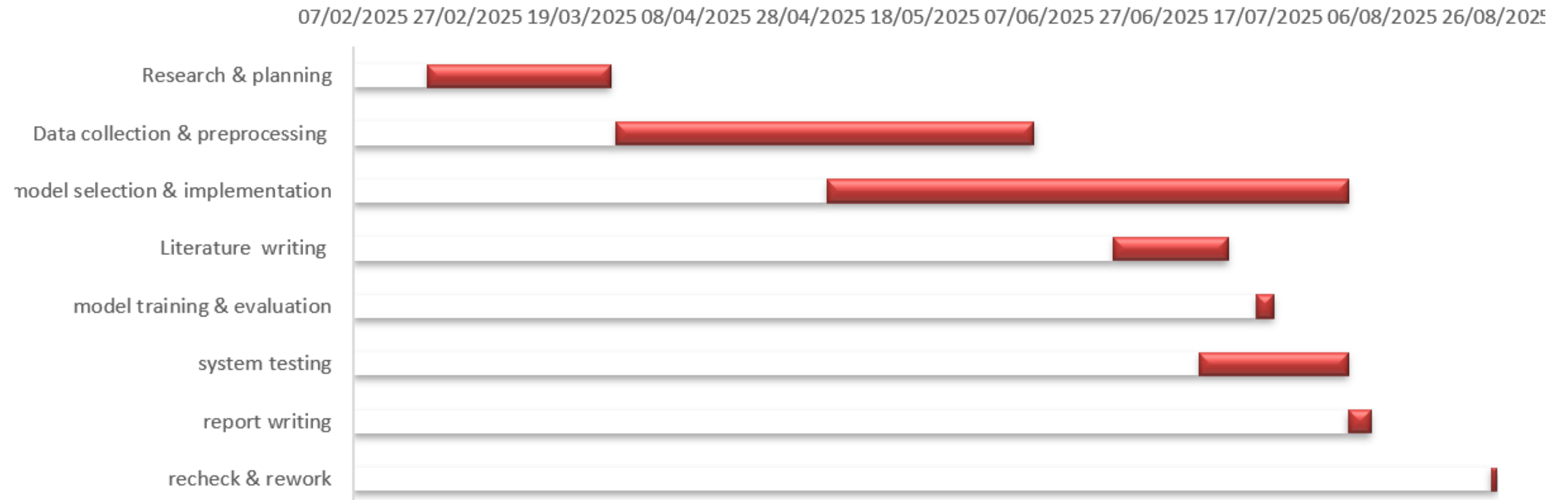
Evaluation Metrics: Need to perform on the evaluation metrics. These include the precision and recall and F1-score and ROC-AUC for validating accurate outbreak predictions and model performance.

Methodological Framework: The famous CRISP-DM framework will be considered for the project work.

Project Planning and Resources

- **Timeline:** The create project plan will be followed here for completing the work with progressive way.
- **Resources Required:** This project needs four primary resources which include specialized computing capacity together with various healthcare information sources alongside sophisticated software platforms for machine learning deployments [5].
- **Risk Management:** The project uses selected datasets for data bias reduction and mitigates model overfitting through regular evaluation and optimization of the system to ensure reliable and robust performance.

Risk	Likelihood	Impact	Mitigation Strategy
Data Privacy Issues	High	High	Implement strict data encryption and access controls. Comply with healthcare regulations like HIPAA.
Inconsistent Data Quality	Medium	High	Use data preprocessing techniques and ensure diverse, high-quality data sources.
Model Overfitting	Medium	Medium	Regular model validation, cross-validation techniques, and data augmentation.
Limited Real-Time Data Access	High	High	Establish partnerships with healthcare providers for live data integration.
Computational Resource Constraints	Medium	Medium	Utilize cloud-based computing solutions and optimize model efficiency.

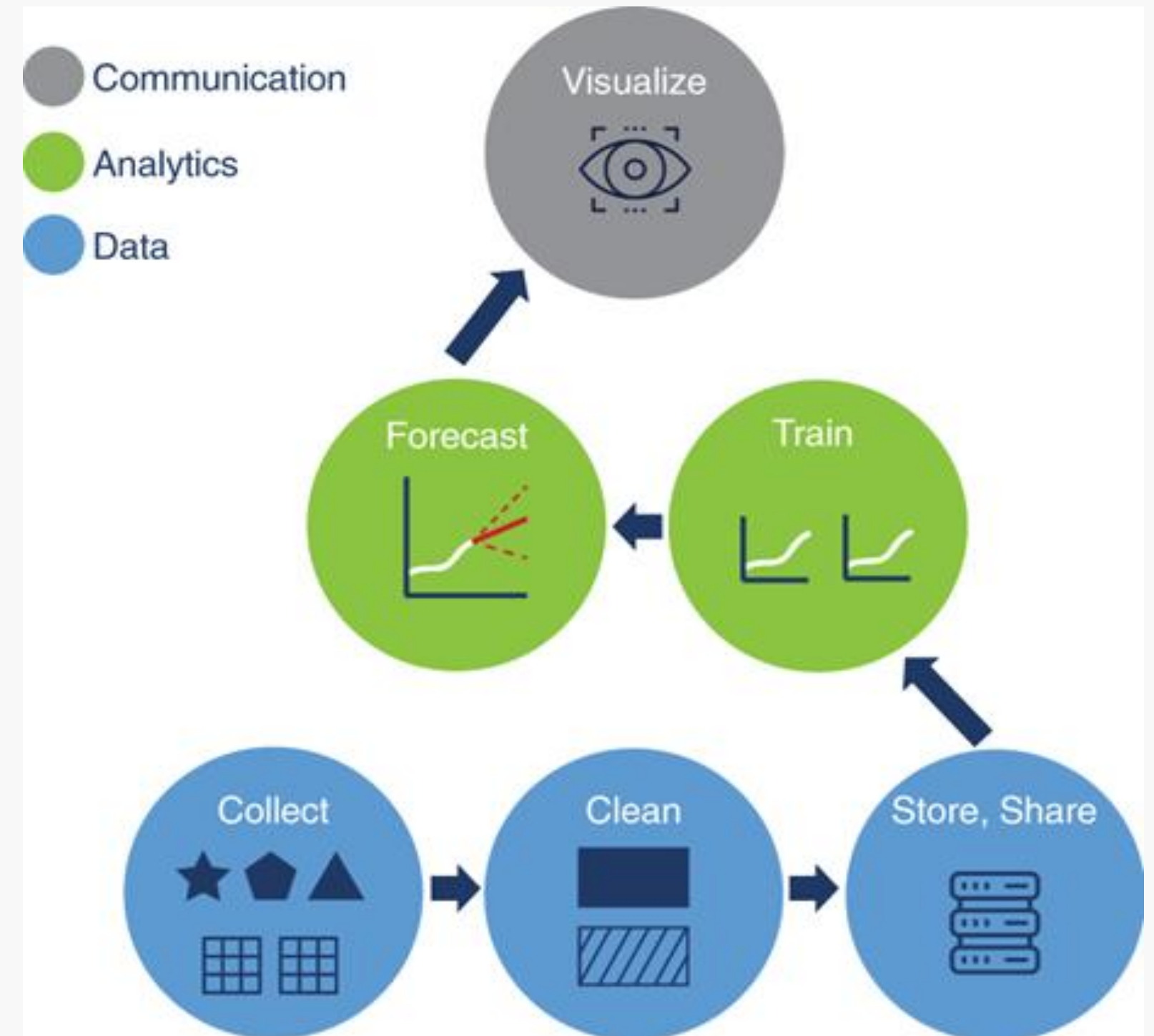


Project Planning and Resources

Project Impact and Limitations



- Potential Benefits: Early detection of disease outbreaks produces benefits where healthcare institutions can implement quick interventions to reduce health system costs as well as enhance public health responses.
- Limitations: The sensitive information present in healthcare records raises privacy issues that create limitations. Model reliability suffers when healthcare professionals experience difficulty obtaining real-time data of high quality [6].
- Future Scope: Future developments will focus on developing adaptable ML models for real-time data streams which will boost outbreak prediction accuracy within healthcare platforms.



Conclusion and Next Steps



Summary:

The implementation of ML-based outbreak prediction systems helps global healthcare organizations detect threats in advance and implement preventive measures.

Implementation Strategy:

The developed model will be installed within healthcare platforms to process current data effectively for accurate prediction outputs. Working hand in hand with healthcare groups enables effectiveness.



Future Research:

Future studies should focus on building larger multi-source data collections along with developing better deep learning algorithms and making models more flexible to enhance both prediction accuracy and real-time outbreak surveillance ability.



References

- [1] Igwama, G.T., Olaboye, J.A., Maha, C.C., Ajegbile, M.D. and Abdul, S., 2024. Big data analytics for epidemic forecasting: Policy Frameworks and technical approaches. *International Journal of Applied Research in Social Sciences*, 6(7), pp.1449-1460. DOI: 10.51594/ijarss.v6i7.1334
 - [2] Santangelo, O.E., Gentile, V., Pizzo, S., Giordano, D. and Cedrone, F., 2023. Machine learning and prediction of infectious diseases: a systematic review. *Machine Learning and Knowledge Extraction*, 5(1), pp.175-198. <https://doi.org/10.3390/make5010013>.
 - [3] Leung, X.Y., Islam, R.M., Adhami, M., Ilic, D., McDonald, L., Palawaththa, S., Diug, B., Munshi, S.U. and Karim, M.N., 2023. A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLoS neglected tropical diseases*, 17(2), p.e0010631. <https://doi.org/10.1371/journal.pntd.0010631>
 - [4] Niaz, M. and Nwagwu, U., 2023. Managing healthcare product demand effectively in the post-covid-19 environment: navigating demand variability and forecasting complexities. *American Journal of Economic and Management Business (AJEMB)*, 2(8), pp.316-330. <http://ajemb.us/index.php/gp/article/view/55>
 - [5] Absar, N., Uddin, N., Khandaker, M.U. and Ullah, H., 2022. The efficacy of deep learning-based LSTM model in forecasting the outbreak of contagious diseases. *Infectious Disease Modelling*, 7(1), pp.170-183. <https://doi.org/10.1016/j.idm.2021.12.005>
 - [6] Lim, T.Y., Kanjilal, S., Doron, S., Penney, J., Haddix, M., Koo, T.H., Danza, P., Fisher, R., Grad, Y.H. and Hay, J.A., 2024. Nowcasting epidemic trends using hospital-and community-based virologic test data. *medRxiv*, pp.2024-11. <https://doi.org/10.1101/2024.11.01.24316580>
-

Thank you
